
Aplicación web para visualizar la evolución temporal de una wiki



Trabajo de Fin de Grado
Curso 2018–2019

Autor

Laura de la Fuente Lorenzo
Adrián Burillo Elmaleh
Laura Sánchez de Rojas Huerta

Director

Javier Arroyo Gallardo
Abel Serrano Juste

Grado en Ingeniería Informática
Facultad de Informática
Universidad Complutense de Madrid

Aplicación web para visualizar la evolución temporal de una wiki

Trabajo de Fin de Grado en Ingeniería Informática
Departamento de Ingeniería del Software e Inteligencia Artificial

Autor

Laura de la Fuente Lorenzo
Adrián Burillo Elmaleh
Laura Sánchez de Rojas Huerta

Director

Javier Arroyo Gallardo
Abel Serrano Juste

Convocatoria: *Septiembre 2019*

Calificación: *Nota*

Grado en Ingeniería Informática
Facultad de Informática
Universidad Complutense de Madrid

20 de septiembre de 2019

Autorización de difusión

El abajo firmante, matriculado en el Grado en Ingeniería en Informática de la Facultad de Informática, autoriza a la Universidad Complutense de Madrid (UCM) a difundir y utilizar con fines académicos, no comerciales y mencionando expresamente a su autor el presente Trabajo Fin de Grado: “Aplicación Web para visualizar la evolución temporal de una wiki”, realizado durante el curso académico 2018-2019 bajo la dirección de Javier Arroyo Gallardo y Abel Serrano Juste en el Departamento de Ingeniería del Software e Inteligencia Artificial, y a la Biblioteca de la UCM a depositarlo en el Archivo Institucional E-Prints Complutense con el objeto de incrementar la difusión, uso e impacto del trabajo en Internet y garantizar su preservación y acceso a largo plazo.

Laura de la Fuente Lorenzo, Adrián Burillo Elmaleh y Laura Sánchez
de Rojas Huerta

20 de septiembre de 2019

Agradecimientos

En primer lugar, queremos agradecer a nuestros directores del TFG por proporcionarnos las guías adecuadas cuando lo hemos necesitado y por acogernos como tutores. En concreto, a Abel Serrano por el desarrollo de WikiChron Classic, la aplicación previa a WikiChron Monowiki, ya que sin ella, este trabajo no hubiera sido posible, y por la creación de múltiples scripts y el parser que facilita la obtención de los datos del dump de Wikia, en el que nos hemos basado para nuestro trabajo en la línea en métricas centradas en el estudio del conocimiento. Además, agradecerle su asistencia a la hora de unificar WikiChron Monowiki con las otras dos versiones de WikiChron (WikiChron Classic y Networks). A Javier Arroyo por su conocimiento a la hora de ayudarnos a clarificar en qué dirección dirigir nuestro trabajo, a determinar qué métricas implementar en WikiChron Monowiki, y a ponernos sobre aviso acerca de la existencia de herramientas que han hecho nuestro trabajo más sencillo en la línea de trabajo de métricas centradas en el estudio del conocimiento, como la librería NLTK de Python.

También queremos agradecer a nuestra familia, amigos, pareja y compañeros de clase por su apoyo e interés durante el desarrollo de este trabajo.

Por último, agradecer a la Universidad Complutense de Madrid y a sus profesores por guiarnos en la adquisición de todos los conocimientos que hemos empleado en el desarrollo de este proyecto, y sin los que no hubiera sido posible este trabajo.

Resumen

Desde sus inicios, internet ha sido una herramienta para compartir conocimiento. Durante su más temprano auge, en la década de los 90, la web estaba formada por el conocimiento aportado por unos pocos colaboradores con unas habilidades de software que todavía no estaban tan extendidas. A día de hoy, las barreras que entonces impedían el acceso y la contribución en la web, han desaparecido (Warschauer, 2007), con lo que se ha producido el nacimiento de la web 2.0, que ha pasado de ser una “web de leer” a una “web de leer y escribir” (Ullrich et al., 2008). Así, surge la creación de proyectos colaborativos, que actualmente, atraen a múltiples usuarios.

Existen gran cantidad de proyectos colaborativos exitosos, como Wikipedia, Open Street Map o Arduino. No obstante, con este trabajo de fin de grado, pretendemos completar los escasos estudios e hipótesis ya existentes sobre proyectos con un éxito menos rotundo, o incluso fallidos, a los que los estudios existentes no han prestado suficiente atención: repositorios de wikis, entre los que destaca Wikia, también conocida como Fandom.

Para ello, nuestro trabajo se basa en colaborar en la mejora y crecimiento de una aplicación web ya existente, WikiChron Classic, que permite analizar la evolución de múltiples wikis simultáneamente, ampliando sus funcionalidades para mejorar el estudio de estas comunidades, y creando otra versión de la herramienta: WikiChron Monowiki. Esta nueva versión está orientada a llevar a cabo un estudio más exhaustivo de una única wiki, permitiendo visualizar en una misma gráfica, diversos factores de la misma, que permiten observar en mayor profundidad qué sucede en la comunidad en un período de tiempo concreto.

Para ello, WikiChron Monowiki hace uso de gráficos más detallados, que permite visualizar diversos factores de forma simultánea. Estos gráficos permiten observar la evolución temporal de aspectos como la composición de la comunidad, atendiendo a distintas categorías de usuarios, o la actividad de los usuarios en distintas partes de la wiki. En definitiva, los gráficos temporales de WikiChron Monowiki permiten estudiar este tipo de comunidades en

un nivel más profundo y detallado que en la versión previa de la herramienta.

Para ello, en este trabajo hacemos uso de algunas tecnologías muy emergentes, como Python, un lenguaje que proporciona multitud de posibilidades a la hora de crear gráficos que permitan la visualización y el análisis de datos, y en especial sus librería Pandas, Plotly y la librería NLTK.

Palabras clave

Comunidades colaborativas, Wikia, aplicación web, visualización de datos, análisis de datos, Python.

Abstract

Ever since the beginning, the Internet has been a knowledge tool for the human race. During its early boom, in the 90's decade, the Web was built up mostly by the knowledge of a few collaborators, who held software skills which weren't as spread as they are today.

However, nowadays, the existing barriers which kept common users from accessing and contributing in the Web, are gone (Warschauer, 2007). This growth has led to the rise of what is known as the Web 2.0, which has evolved from just a “reading Web” to a “reading and writing Web” (Ullrich et al., 2008). All along with this event, collaborative communities have risen.

There are several successful collaborative projects in the air, like Wikipedia, Open Street Map or Arduino. Nevertheless, in this Degree final project, we focus on extending and completing the few existing studies and literature about less successful or even failed communities, to which existing literature and research hasn't paid enough attention: wiki repositories, paying special attention to Wikia, also known as Fandom.

In order to do so, our Degree final project is based on collaborating on the improvement and growth of an existing tool, WikiChron Classic, which allows to simultaneously analyze multiple wikis, contributing to enlarge its functionalities to better understand these communities dynamics. Also, we will create another version of the tool, WikiChron Monowiki, which will be focused on allowing a deeper study of a single wiki, noticing aspects such as the underlying community structure.

In order to achieve it, WikiChron Monowiki includes more detailed visualizations, which allow to observe the temporal evolution of aspects such as the community composition in terms of various user categories, or the activity of editors in different parts of the wiki. Definitely, the temporal series developed in WikiChron Monowiki allow to analyze this type of communities more deeply than on its previous version.

In order to do so, in this Degree final project, we use some of the most emerging technologies, such as python, a programming language which pro-

vides multiple functionalities and alternatives to data analysis, and particularly, its libraries: Pandas, Plotly and NLTK.

Keywords

Collaborative communities, Wikia, Web tool, data analysis, data visualization, Python.

Índice

1. Introducción	1
1.1. Contexto	1
1.1.1. El auge de las comunidades colaborativas	1
1.1.2. La visualización de datos como herramienta para extraer conocimiento	2
1.2. Motivaciones	3
1.3. Objetivos	4
1.3.1. Investigación y estudio de las comunidades colaborativas	4
1.3.2. Decidir aspectos importantes y mejorar su visualización	5
1.3.3. WikiChron Monowiki	5
1.4. Estructura del documento	5
1. Introduction	7
1.1. Context	7
1.1.1. The rise of collaborative communities	7
1.1.2. Data visualization and progress	8
1.2. Motivations	8
1.3. Objectives	9
1.3.1. Study and research in collaborative communities area	9
1.3.2. Deciding important aspects and improve their visualization	10
1.3.3. WikiChron Monowiki	10
1.4. Document structure	10
2. Marco teórico	13
2.1. El modelo de producción colaborativa	13
2.2. La tecnología Wiki	16
2.3. Wikipedia	17
2.3.1. Estructura de la comunidad y gobernanza	18

2.3.2.	Conflictos en la comunidad	20
2.3.3.	Participación	23
2.3.4.	Evolución del contenido	24
2.4.	Fandom	28
2.4.1.	Estructura de la comunidad	28
2.4.2.	Participación	31
3.	Herramientas	33
3.1.	Herramientas genéricas para visualización de datos	33
3.1.1.	Tableau Software	34
3.1.2.	Google Analytics	34
3.1.3.	Conclusiones	36
3.2.	Herramientas centradas en Software Libre	36
3.2.1.	Openhub	36
3.3.	Bitergia Cauldron	37
3.3.1.	Conclusiones	38
3.4.	Herramientas específicas para el estudio de wikis	39
3.4.1.	WikiStats	39
3.4.2.	WAM de Wikia	41
3.4.3.	WikiApiary	45
3.4.4.	Conclusiones	45
3.5.	WikiChron	46
3.5.1.	Tecnologías.	46
3.5.2.	Formato de datos	47
3.5.3.	Estructura de la APP	48
4.	Contribuciones en WikiChron Classic	57
4.1.	Usuarios activos registrados y anónimos	57
4.2.	Usuarios activos con más de 4, 24 o 99 ediciones	58
4.3.	Wiki durmiente	59
4.4.	Métricas de retención de usuarios de Wikimedia	60
4.4.1.	Nuevo editor reincidente	60
4.4.2.	Nuevo editor superviviente	61
5.	WikiChron Monowiki	63
5.1.	Motivaciones	63
5.2.	Visualización	64
5.2.1.	Nuevos gráficos utilizados	64
5.2.2.	Modificaciones para incluir los nuevos gráficos	65
5.3.	Métricas centradas en la distribución de usuarios registrados activos	66
5.3.1.	Usuarios activos según experiencia editando	68

5.3.2.	Usuarios activos por antigüedad	69
5.3.3.	Usuarios activos por retorno	70
5.3.4.	Usuarios activos por racha de actividad	71
5.3.5.	Usuarios activos en función del tipo de página editada	74
5.3.6.	Ediciones en diferentes tipos de páginas	76
5.4.	Métricas centradas en la distribución de ediciones según tipos de usuarios registrados activos	77
5.5.	Métricas centradas en las ediciones	82
5.5.1.	Porcentaje de editores que han hecho un porcentaje fijado de ediciones	82
5.5.2.	Distribución de ediciones entre usuarios registrados . .	84
5.5.3.	Bytes añadidos y eliminados en los artículos	85
5.5.4.	Distribución de usuarios registrados según el número de artículos editados mensualmente.	87
5.5.5.	Distribución de las ediciones en páginas	88
5.5.6.	Distribución de revisiones en páginas	88
5.6.	Métricas centradas en unidades de conocimiento	90
5.6.1.	Acceso al contenido de las ediciones del dump de Wikia	91
5.6.2.	Conteo de factoides añadidos y borrados	91
5.6.3.	Distribución de factoides añadidos y borrados por tipos de usuario	94
5.7.	Métricas centradas en enlaces e imágenes en artículos	100
6.	Análisis de wikis	103
6.1.	Introducción	103
6.2.	Tendencias estructurales opuestas en wikis de hobbies	103
6.2.1.	Cocktails	105
6.2.2.	Home	109
6.2.3.	Conclusiones	113
6.3.	Análisis del compromiso de los usuarios en distintas wikis . .	115
6.3.1.	Laguna negra (El Internado)	116
6.3.2.	WikiAnswer Pets	125
6.4.	Análisis sobre ediciones	130
6.4.1.	Análisis con la wiki de Lego(ES)	130
6.4.2.	Análisis con la wiki de Juego de Tronos(GoT)	134
6.5.	Análisis con Factoides	139
7.	Conclusiones y Trabajo Futuro	145
7.1.	Conclusiones	145
7.2.	Trabajo futuro	147
7.2.1.	Visualizar la evolución del texto de los artículos a lo largo del tiempo	147

7.2.2.	Refinar la categorización gramatical de las palabras en artículos	148
7.2.3.	Añadir información sobre la intencionalidad de las ediciones utilizando un clasificador	148
7.	Conclusions and Future Work	149
7.1.	Conclusions	149
7.2.	Future work	150
7.2.1.	Visualizing text evolution in articles across time	151
7.2.2.	Improve part of speech tagging in article words	151
7.2.3.	Adding intentions information to edits	152
8.	Contribuciones individuales	153
8.1.	Laura Sánchez de Rojas Huerta	153
8.2.	Laura De La Fuente Lorenzo	154
8.3.	Adrián Burillo Elmaleh	156
	Bibliografía	159

Índice de figuras

2.1. Regla del 90-9-1.	15
2.2. Lista de vigilancia de Wikipedia	21
3.1. hoja de trabajo en Tableau	35
3.2. Información general de los proyectos a comparar en ohloh.net	37
3.3. Estadísticas generales de los proyectos a comparar en ohloh.net	37
3.4. Estadísticas generales de código en los proyectos a comparar en ohloh.net	38
3.5. Reporte de actividad de la herramienta Cauldron desarrollada por Bitergia.	38
3.6. Métrica de aparatos únicos conectados en WikiStats.	39
3.7. Métrica de número total de páginas visitadas en WikiStats. .	40
3.8. Métrica de número total de páginas visitadas por país en Wi- kiStats.	40
3.9. Métrica de ediciones en WikiStats.	41
3.10. Métrica de usuarios nuevos registrados en WikiStats.	42
3.11. Métrica de usuarios nuevos registrados en WikiStats.	42
3.12. Métrica de páginas editadas en WikiStats.	43
3.13. Métrica de diferencia neta en bytes en WikiStats.	43
3.14. Métrica de diferencia absoluta en bytes en WikiStats.	44
3.15. WAM de Wikia.	44
3.16. Active Users	46
3.17. Página principal WikiChron V1	48
3.18. Gráfico con una wiki seleccionada	49
3.19. Gráfico acortando el date slider y seleccionando meses desde el nacimiento	49
3.20. Selección de varias 2 métricas y 2 wikis	50
3.21. Diagrama Wikichron v1	51
3.22. Página principal WikiChron V2	53

3.23. Pantalla de selección de Wikis.	53
3.24. Pantalla de selección de métricas.	54
3.25. Pantalla donde aparecen los gráficos.	54
3.26. Diagrama estructura WikiChron V2	55
4.1. Métrica de usuarios activos anónimos, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012. .	58
4.2. Métrica de usuarios activos registrados, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.	58
4.3. Métrica de usuarios activos con más de 4 ediciones, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.	58
4.4. Métrica de usuarios activos con más de 24 ediciones, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.	59
4.5. Métrica de usuarios activos con más de 99 ediciones, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.	59
4.6. Métrica de wiki durmiente, en la wiki de Cocktails, durante el periodo temporal de abril de 2005 a julio de 2012.	60
4.7. Métrica de Returning new editor, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012. .	61
4.8. Métrica de Surviving new editor, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012. . .	62
5.1. Generación de gráficos: antes y después de las modificaciones introducidas para la inclusión de distintos modos de visualización.	65
5.2. Métrica de distribución de usuarios activos registrados según su experiencia editando en la wiki de Cocktails, en el periodo temporal desde abril de 2008 hasta julio de 2012.	69
5.3. Métrica de distribución de usuarios activos registrados según su antigüedad en la wiki de Cocktails, desde abril de 2008 hasta julio de 2012.	70
5.4. Métrica de distribución de usuarios activos registrados por retorno en la wiki de Cocktails, durante el período de abril de 2008 hasta julio de 2012.	71
5.5. Métrica de distribución de usuarios activos registrados por racha de actividad, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.	72

5.6. Métrica de distribución de usuarios activos registrados por racha de actividad en artículos, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.	73
5.7. Métrica de distribución de usuarios activos registrados por su experiencia editando (versión absoluto) en la wiki de Cocktails, durante el período de abril de 2008 hasta julio de 2012.	74
5.8. Métrica de usuarios activos registrados por ediciones según el tipo de página, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.	76
5.9. Métrica de usuarios activos registrados por ediciones según el resto de tipos de página, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.	77
5.10. Métrica de distribución de ediciones por usuarios activos según su experiencia editando, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.	78
5.11. Métrica de distribución de ediciones por usuarios activos según su antigüedad, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.	79
5.12. Métrica de distribución de ediciones por usuarios activos según su última edición, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.	79
5.13. Métrica de distribución de ediciones por usuarios activos según su racha actual, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.	80
5.14. Métrica de distribución de ediciones por usuarios activos según su racha actual en artículos, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.	80
5.15. Métrica de distribución de ediciones por usuarios activos según el tipo de página editada, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.	81
5.16. Métrica de distribución de ediciones por usuarios activos según el resto tipos de página editada, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.	81
5.17. Métrica de porcentaje de editores por el porcentaje de ediciones (versión acumulada) en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.	83
5.18. Métrica de porcentaje de editores por el porcentaje de ediciones (versión mensual) en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.	84
5.19. Métrica de distribución de usuarios entre ediciones, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.	85

5.20. Métrica de bytes añadidos entre artículos, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.	86
5.21. Métrica de bytes eliminados entre artículos, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.	86
5.22. Métrica de distribución de usuarios registrados entre artículos por mes, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.	87
5.23. Métrica de ediciones en páginas, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.	88
5.24. Métrica de revisiones en páginas, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.	89
5.25. Ejemplo gráfico de factoids añadidos en distintas revisiones de varias páginas. Nótese que de una revisión a otra, pueden añadirse y borrarse factoids	92
5.26. Métrica de Added factoids en la wiki de Cocktails, en el periodo temporal de abril de 2008 a julio de 2012.	92
5.27. Métrica de Total added factoids en la wiki de Cocktails, en el periodo temporal de abril de 2008 a julio de 2012.	93
5.28. Ejemplo gráfico de factoids borrados en distintas revisiones de una misma página.	93
5.29. Métrica de Deleted factoids en la wiki de Cocktails en el periodo temporal de abril de 2008 a julio de 2012.	94
5.30. Métrica de Total deleted factoids en la wiki de Cocktails en el periodo temporal de abril de 2008 a julio de 2012.	94
5.31. Métrica de Added factoids by active users by editing experience en la wiki de Cocktails en el periodo temporal de abril de 2008 a julio de 2012.	96
5.32. Métrica de Added factoids by active users by tenure en la wiki de Cocktails en el periodo temporal de abril de 2008 a julio de 2012.	96
5.33. Métrica de Added factoids by users by edit streak en la wiki de Cocktails en el periodo temporal de abril de 2008 a julio de 2012.	97
5.34. Métrica de Added factoids by active users by the date of the last edit en la wiki de Cocktails en el periodo temporal de abril de 2008 a julio de 2012.	97
5.35. Métrica de Deleted factoids by active users by editing experience en la wiki de Cocktails en el periodo temporal de abril de 2008 a julio de 2012.	98

5.36. Métrica de Deleted factoids by active users by tenure en la wiki de Cocktails en el período temporal de abril de 2008 a julio de 2012.	99
5.37. Métrica de Deleted factoids by active users by edit streak en la wiki de Cocktails en el período temporal de abril de 2008 a julio de 2012.	99
5.38. Métrica de Deleted factoids by active users by the date of the last edit en la wiki de Cocktails en el período temporal de abril de 2008 a julio de 2012.	100
5.39. Métrica de conteo de enlaces, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.	101
5.40. Métrica de conteo de imagenes, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012. . .	101
5.41. Métrica que muestra la división entre enlaces externos e internos, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.	101
6.1. Métrica de Active registered users by experience (in percentage), en la wiki de Cocktails, durante sus primeros 120 meses.	105
6.2. Métrica de Edits by editing experience (in percentage), en la wiki de Cocktails, durante sus primeros 120 meses.	105
6.3. Métrica de Active registered users by experience, en la wiki de Cocktails, durante sus primeros 120 meses.	106
6.4. Métrica de Edits by editing experience, en la wiki de Cocktails, durante sus primeros 120 meses.	106
6.5. Métrica de Active registered users by tenure (in percentage), en la wiki de Cocktails, durante sus primeros 120 meses. . . .	107
6.6. Métrica de Active registered users by the date of the last edit (in percentage), en la wiki de Cocktails, durante sus primeros 120 meses.	107
6.7. Métrica de Active registered users by editing experience (in percentage), en la wiki de Home, durante sus primeros 120 meses.	109
6.8. Métrica de Edits by editing experience (in percentage), en la wiki de Home, durante sus primeros 120 meses.	110
6.9. Métrica de Active registered users by experience, en la wiki de Home, durante sus primeros 120 meses.	110
6.10. Métrica de Edits by editing experience, en la wiki de Home, durante sus primeros 120 meses.	111
6.11. Métrica de Active registered users by tenure (in percentage), en la wiki de Home, durante sus primeros 120 meses.	111

6.12. Métrica de Active registered users by the date of the last edit (in percentage), en la wiki de Home, durante sus primeros 120 meses.	112
6.13. Métrica de Total edits in pages de WikiChron Classic, en las wikis de Home (naranja) y Cocktails (azul), durante sus primeros 120 meses.	113
6.14. Métrica del Ratio 10:90, en las wikis de Home (naranja) y Cocktails (azul), durante sus primeros 120 meses.	114
6.15. Métrica de usuarios activos registrados según su última edición, en la wiki de Laguna negra, desde abril de 2009 hasta mayo de 2010	116
6.16. Métrica de usuarios activos registrados según su racha actual, en la wiki de Laguna negra, desde abril de 2009 hasta abril de 2010	117
6.17. Métrica de usuarios activos registrados según su racha actual, en la wiki de Laguna negra, desde mayo de 2009 hasta noviembre de 2010	118
6.18. Métrica de nuevo usuario reincidente, en la wiki de Laguna negra, desde abril de 2009 hasta enero de 2011	118
6.19. Métrica de usuario activos registrados según su última edición, en la wiki de Laguna negra, desde abril de 2009 hasta abril de 2011	119
6.20. Métrica de usuarios activos registrados según su racha actual, en la wiki de Laguna negra, desde abril de 2009 hasta febrero de 2012	120
6.21. Métrica de usuarios activos registrados según su racha actual, en la wiki de Laguna negra, desde mayo de 2012 hasta septiembre de 2014	121
6.22. Métrica de usuarios activos registrados según su última edición, en la wiki de Laguna negra, desde octubre de 2016 hasta enero de 2019	122
6.23. Métrica de usuarios activos registrados según su racha actual, en la wiki de Laguna negra, desde julio de 2016 hasta enero de 2019	122
6.24. Métrica de dist. ediciones por usuarios activos registrados según su antigüedad, en la wiki de Laguna negra, desde abril de 2009 hasta abril de 2012	123
6.25. Métrica de dist. ediciones por usuarios activos registrados según su racha actual, en la wiki de Laguna negra, desde abril de 2009 hasta abril de 2012	124
6.26. Métrica de dist. ediciones por usuarios activos registrados según su última edición, en la wiki de Laguna negra, desde abril de 2009 hasta abril de 2012	124

6.27. Métrica de usuarios activos registrados según su última edición, en la wiki de Wikianswer Pets, desde junio de 2007 hasta julio de 2010	125
6.28. Métrica de nuevo usuario reincidente, en la wiki de Wikianswer Pets, desde julio de 2007 hasta enero de 2011	126
6.29. Métrica de usuarios activos registrados según su racha actual, en la wiki de Wikianswer Pets, desde julio de 2007 hasta octubre de 2010	126
6.30. Métrica de usuarios activos registrados según su antigüedad, en la wiki de Wikianswer Pets, desde julio de 2007 hasta enero de 2011	127
6.31. Métrica de usuarios activos registrados según su antigüedad, en la wiki de Wikianswer Pets, desde enero de 2011 hasta enero de 2019	127
6.32. Métrica de usuarios activos registrados según su antigüedad, en la wiki de Wikianswer Pets, desde julio de 2007 hasta enero de 2012	128
6.33. Métrica de usuarios activos registrados según su racha actual, en la wiki de Wikianswer Pets, desde julio de 2007 hasta enero de 2012	129
6.34. Métrica de usuarios activos registrados según su última edición, en la wiki de Wikianswer Pets, desde julio de 2007 hasta enero de 2012	129
6.35. Lego(Es) : Datos de la wiki	130
6.36. Métrica de dist. ediciones entre usuarios activos registrados según el tipo de página editada, en la wiki de Lego	130
6.37. Métrica de dist. ediciones entre usuarios activos registrados según el tipo de página editada tomando las categorías de menos importancia, en la wiki de Lego	131
6.38. Métrica de dist. ediciones en páginas, en la wiki de Lego	132
6.39. Métrica de dist. revisiones en páginas, en la wiki de Lego	132
6.40. Métrica de dist. editores entre artículos, en la wiki de Lego	133
6.41. Datos de GOT	135
6.42. Métrica de dist. ediciones entre usuarios activos registrados según el tipo de página editada, en la wiki de GOT	135
6.43. Métrica de dist. ediciones entre usuarios activos registrados según el tipo de página editada tomando las categorías de menos importancia, en la wiki de GOT	136
6.44. Métrica de dist. ediciones en páginas, en la wiki de GOT	137
6.45. Métrica de dist. revisiones en páginas, en la wiki de GOT	137
6.46. Métrica de dist. editores entre artículos, en la wiki de GOT	138
6.47. Datos de World Witches Series	140

6.48. Métrica de número de factoides añadidos por editores según su experiencia editando, en la wiki de Word Witches Series	140
6.49. Métrica de número de factoides eliminados por editores según su primera edición, en la wiki de Word Witches Series	141
6.50. Métrica de número de factoides añadidos por editores según su racha actual, en la wiki de Word Witches Series	142
6.51. Métrica de número de factoides eliminados por editores según su última edición, en la wiki de Word Witches Series	143
7.1. Visualización de la interfaz de la herramienta History Flow para la página Microsoft de Wikipedia. A la izquierda, se muestra el contenido de la página, a la derecha, todos los autores que la han editado con un color asignado, y en el centro, la evolución del contenido de la misma.	147
7.1. History flow visualization for the Wikipedia Microsoft page. We can see the content of the page at right, all the authors that have contributed to it at left, and the visualization of the evolution of the article content in the center.	151

Capítulo 1

Introducción

En este capítulo, se explica, en primer lugar, el contexto en el que se sitúan los factores en torno a los que gira nuestro TFG, siendo estos: el auge de los proyectos colaborativos en nuestra sociedad y de la visualización de datos como técnica indispensable para el análisis científico y el progreso en diversas áreas de nuestra sociedad. Posteriormente, en base a este contexto, se exponen las motivaciones que alimentan nuestro proyecto, siguiendo con los objetivos del mismo y terminando con la estructura del documento.

1.1. Contexto

1.1.1. El auge de las comunidades colaborativas

Los proyectos o comunidades colaborativas componen un sistema socio-económico basado en compartir y consumir información que está proliferando en las plataformas digitales. La implicación en este tipo de proyectos es considerada como un fenómeno muy positivo por la literatura actual, ya que favorece el desarrollo de determinadas características del ser humano, como el voluntariado, o la autonomía, permitiéndole transitar de un estado en el que simplemente recibe información, a un estado en el que puede participar activamente en ella (Benkler y Nissenbaum, 2006). Desde el nacimiento de la Web 2.0, momento en el que desaparecieron las barreras económicas y tecnológicas que impedían el acceso a la Web, la participación y creación de nuevos proyectos colaborativos ha proliferado notablemente (Cummings, 2006).

Se puede afirmar que los valores positivos que promueven los proyectos colaborativos en el ser humano, y la facilidad para crearlos y participar en ellos, son dos de los factores más influyentes en la proliferación de este tipo de comunidades. Las características principales de este tipo de comunidades son que se basan en la producción voluntaria a través de plataformas digitales,

y están compuestas por relaciones de “pares”, es decir, entre iguales. En comunidades colaborativas, este término hace referencia la organización del trabajo descentralizada donde en principio todos los miembros son iguales. Estas relaciones de pares no se basan en ningún contrato ni en ningún fin comercial, sino que los individuos son totalmente autónomos (Fuster Morell et al., 2015).

Cabe destacar que cuando hablamos de proyectos colaborativos de producción de bienes comunes, hacemos referencia a la definición proporcionada por Bates (2007): proyectos llevados a cabo de manera altamente distribuida, y cuyo objetivo es la producción de bienes comunes. Esto incluye a todo tipo de comunidades en las que se genera un producto a través la participación voluntaria de un grupo de personas, ya sea de carácter meramente informativo, como Wikipedia, software libre, como Linux, repositorios online como Github, productos basados en hardware como Arduino, etcétera.

Estos proyectos generan un gran valor y la forma en que se producen es muy innovadora y por eso está siendo una disciplina muy estudiada desde distintos frentes: sociología, economía, informática, ciencia de datos, etc.

En estas comunidades el bien producido se almacena en la nube. Además, estos repositorios en muchos casos guardan un histórico que refleja todas las interacciones entre los colaboradores y los distintos estados que va atravesando dicho bien común. Esto es una fuente de datos muy rica para intentar entender los procesos de organización, los problemas que sufren estas comunidades, la razón por la que unos proyectos son exitosos y muchos otros no, etcétera.

1.1.2. La visualización de datos como herramienta para extraer conocimiento

Vivimos en una sociedad que se encuentra en continua mejora, y para conseguirla, una de las técnicas más empleadas es la visualización y el análisis de datos, que hace de la toma de decisiones, un proceso mejor documentado, y aumenta la probabilidad de éxito de la decisión tomada. La visualización de datos es el proceso de adquisición, limpieza y transformación de los mismos en información útil para el usuario. Cabe destacar que antes de poder comunicar la información de forma efectiva al usuario final, los encargados de la visualización han de conocer en profundidad qué información están transmitiendo y sus características.

Durante la transición hasta la Web 2.0, hay un crecimiento de la abundancia de datos, y por tanto, la tarea de extraer información de los mismos es cada vez más compleja. Para resolver este problema, existen numerosas herramientas de visualización, como por ejemplo, Tableau y Google Analytics. Este tipo de herramientas supone un arma fundamental para los analistas, que bien orientada hacia el fin que se pretende alcanzar, facilita la reten-

ción de la información, y la extracción de los datos cuya importancia es más relevante para las decisiones o hipótesis que se pretenden demostrar.

Así, las empresas desarrolladoras de software, han invertido grandes cantidades de dinero en la implementación de estas herramientas, que en numerosas ocasiones, están orientadas a un objetivo o área de estudio concreta. Se puede afirmar que la tarea del analista de datos reside también en saber identificar el fin original de la herramienta, para que ésta no se convierta en un obstáculo que entorpezca su análisis.

Por otra parte, la visualización de datos en el área del estudio de wikis, se encuentra en necesidad de progreso, dado que, como mencionaremos en secciones posteriores de este documento, hay pocas herramientas de libre acceso especialmente orientadas al estudio de este área, por lo que la tarea de los analistas no sólo reside en sacar conclusiones de lo visualizado, sino también en crear la propia herramienta.

1.2. Motivaciones

Situándonos en el contexto anteriormente expuesto, comenzando con el auge y la proliferación de las comunidades colaborativas, pretendemos, en primer lugar, proporcionar al lector un trasfondo teórico que muestre cuáles son los proyectos colaborativos más populares en la actualidad, sus características, estudios e hipótesis acerca de los mismos. A pesar de dotar al lector de este conocimiento sobre comunidades colaborativas muy exitosas como Wikipedia, en este trabajo de fin de grado, hacemos hincapié en la necesidad de centrarnos en el análisis de proyectos colaborativos en distintas etapas de crecimiento y madurez, incluyendo aquellos que se estancan. Para ello, nuestra motivación es la creación de una herramienta que permita mostrar de forma objetiva qué sucede en una comunidad a lo largo del tiempo, y que permita entender cómo evoluciona la comunidad y la participación en la misma de sus distintos miembros. Por otra parte, también es interesante que con su uso, se pueda extraer un diagnóstico del estado del proyecto, que puede ser empleado por la propia comunidad y por los estudiosos en el área de las wikis.

Para ello, pretendemos ampliar la herramienta ya existente, WikiChron Classic (Serrano et al., 2018b), creando una nueva versión de la misma, WikiChron Monowiki, en la que proporcionamos unas definiciones de métricas que permiten desgranar aspectos que quedan ocultos con las métricas ya existentes en la versión previa, y en la que mejoramos la forma de visualizar estas métricas nuevas. Esta nueva versión permitirá observar aspectos más detallados de una única wiki, como la composición de la comunidad de usuarios activos en diversas categorías, o la actividad de los mismos en diversas partes de la wiki.

Además, tal y como se ha mencionado anteriormente, muchos de los es-

tudios existentes, se encuentran incompletos debido a que giran en torno a proyectos muy populares y de éxito rotundo, como por ejemplo el realizado por Yang et al. (2016) o Kittur et al. (2006), que obtienen resultados que tan sólo son aplicables al tipo particular de comunidad estudiado. Es interesante para nosotros continuar y ampliar dichos estudios, empleando WikiChron Monowiki, e incluyendo en ellos los resultados obtenidos para comunidades colaborativas no tan exitosas, además de aportar una explicación más profunda de diagnósticos obtenidos en estudios que sí se centran en comunidades como Fandom, como sucede en el estudio llevado a cabo por Serrano et al. (2018a).

Además, el proyecto se encuentra en un repositorio de Github¹, facilitando la inclusión de contribuciones de otros usuarios, su consecuente enriquecimiento, y el acceso al código para cualquier duda con respecto a lo que se está visualizando.

1.3. Objetivos

Siguiendo la línea del contexto anteriormente expuesto, y las motivaciones subyacentes de nuestro trabajo de fin de grado, los objetivos del mismo son los siguientes:

1.3.1. Investigación y estudio de las comunidades colaborativas

En primer lugar, la investigación, a través de estudios e hipótesis ya existentes, sobre cuáles son los patrones más usuales de jerarquía y comportamiento de los usuarios en las comunidades colaborativas en general, para conocerlos mejor y facilitar nuestra tarea a la hora de ponernos en la piel del analista. Llevaremos a cabo este proceso de familiarización recopilando estudios ya existentes sobre este tipo de comunidades, como el de Serrano et al. (2018a), en el que se usa la versión de WikiChron Classic para estudiar varios fenómenos en wikis de Fandom, o estudios que se centran en otro tipo de comunidades, como los anteriormente mencionados, de Kittur et al. (2006) y Yang et al. (2016), que estudian cómo evoluciona la participación de los usuarios en Wikipedia y Open Street Map, respectivamente, o el realizado por Viégas et al. (2004), en el que se estudia la evolución temporal de las páginas, en términos de ediciones y de contenido, de la Wikipedia Inglesa.

¹<https://github.com/wikichron-tfg/WikiChron>

1.3.2. Decidir aspectos importantes y mejorar su visualización

Establecer cuáles de estos aspectos a estudiar en una única Wiki son los más relevantes y cuál es la mejor manera de visualizarlos, en cuanto a tipo de gráfico, para obtener la información principal y omitir irrelevancias. En concreto, elegimos entre cuatro tipos de gráficos: gráficos de línea, mapas de calor o "heatmaps", gráficos de barras y gráficos de área. Cada uno de estos tipos de visualización tiene sus ventajas e inconvenientes a la hora de mostrar los datos, y por ello pretendemos determinar cuál de ellos se ajusta mejor con los aspectos a visualizar. Para ello, después de estudiar cómo se ven los datos con varias visualizaciones, uno de nuestros objetivos es ampliar una funcionalidad ya existente desde WikiChron Classic, permitiendo tener distintos tipos de gráficos en función de la métrica.

1.3.3. WikiChron Monowiki

Crear una nueva versión de la aplicación web existente de WikiChron Classic, llamada WikiChron Monowiki. Ambas versiones tienen la finalidad de posibilitar la generación de un diagnóstico que expone el estado de la comunidad de una wiki en un período de tiempo, para poder emplearlo con diversos fines, como por ejemplo, como diagnóstico para la propia comunidad, o como medio para arrojar luz sobre hipótesis y estudios ya existentes sobre este tipo de comunidades. No obstante, más allá de estos objetivos generales que comparten ambas versiones, mientras que WikiChron Classic se centra en la visualización de fenómenos concretos en varias Wikis simultáneamente, WikiChron Monowiki se centra en el análisis de una única Wiki, permitiendo visualizar aspectos más detallados y profundos sobre la misma, e interrelacionar estos aspectos de la Wiki, con otros de la misma. Por otra parte, el código de la aplicación que une ambas versiones se encuentra en un repositorio de GitHub ², lo que hace el uso de estas herramientas mucho más accesible, además de posibilitar la contribución de personas interesadas en el proyecto.

1.4. Estructura del documento

Para terminar, proporcionamos una breve descripción de en qué consiste cada uno de los capítulos de este documento:

- **El capítulo 2**, proporciona el marco teórico necesario para comprender la historia de las comunidades colaborativas. En concreto, se habla de una comunidad muy exitosa, Wikipedia y de otra conocida pero no tan estudiada, Fandom. Además, se proporciona información que permite

²<https://github.com/wikichron-tfg/WikiChron>

entender y conocer qué estudios e hipótesis existen acerca de ambas comunidades colaborativas.

- **El capítulo 3**, contiene un conjunto de herramientas para el análisis de datos. Se comienza exponiendo algunas herramientas de análisis genéricas, populares en el mundo empresarial, para continuar con una serie de herramientas orientadas al análisis de proyectos colaborativos en general, y terminando con herramientas especialmente orientadas al análisis de wikis, entre ellas, WikiChron Classic, que es el punto de partida de la creación de WikiChron Monowiki.
- **El capítulo 4** proporciona una descripción de nuestras contribuciones en WikiChron Classic.
- **El capítulo 5** proporciona una descripción de WikiChron Monowiki y sus funcionalidades principales, junto con las motivaciones que impulsan su creación, y los cuatro modos de visualización introducidos en la herramienta.
- **El capítulo 6** proporciona una serie de análisis de comunidades de Fandom, llevados a cabo empleando WikiChron Monowiki. Con estos análisis, no solamente demostramos la utilidad de WikiChron Monowiki como herramienta de análisis en wikis, sino que además, ampliamos los estudios e investigaciones existentes acerca de este tipo de comunidades.
- **El capítulo 7** proporciona las conclusiones extraídas de este trabajo de fin de grado, así como el trabajo futuro que puede llevarse a cabo con el objetivo de perfeccionar los resultados obtenidos.
- Por último, **el capítulo 8** contiene una breve descripción de cuál es el aporte en este trabajo de cada uno de los miembros del grupo.

Chapter 1

Introduction

In this chapter, we explain, at first, the context in which the different elements that this Degree final project is related to, are located. This context involves the rise of collaborative projects in our society, and data visualization as an essential technique for scientific analysis. Later in this chapter, the underlying motivations for this project are exposed, followed by the objectives that boost this Degree final project. In the end, for the sake of clarity, the structure of this Degree final project document is explained.

1.1. Context

1.1.1. The rise of collaborative communities

Collaborative communities or projects, are considered, by existing literature, as an element which composes a socioeconomic system based on sharing and consuming information. The influence of this new socioeconomic system is exponentially growing in digital platforms. The involvement in this particular kind of projects is considered by many authors as a positive phenomenon, as it promotes the development of human characteristics, such as volunteering or autonomy, allowing the human being to evolve from a state in which they simply receive information, to a state in which they can actively participate on the creation of this information, Benkler y Nissenbaum (2006). Ever since the beginning of what has been called the Web 2.0, participation and creation of new collaborative projects has exponentially grown, Cummings (2006).

Definitely, we can say that both the positive values promoted by collaborative projects, and the simplicity to create them and to get involved with them, are some of the factors that lead them to their current success.

The main characteristics of this kind of communities are based on the voluntary production of content, through digital platforms. These commu-

nities are composed by peer relationships, i.e. solutions to problems can be developed in a decentralized manner. Moreover, these relationships neither are based on any contract nor have they any commercial purpose, as individuals are completely autonomous and their contributions are voluntary, Fuster Morell et al. (2015).

It should be noted that when we talk about collaborative projects, we are referring to any kind of collaborative community whose ultimate goal is to generate a product by means of the voluntary participation of a group of people. This product can be any kind (i.e. written information, Software, Hardware, etc.), like Wikipedia or Linux.

1.1.2. Data visualization and progress

We live in a society which is deeply committed with its improvement. In order to achieve this social improvement, one of the most emergent techniques is data analysis and visualization. This technique simplifies and increases the success probability of the decision making process.

Data visualization, which emerges hand by hand with the Web 2.0, is the process of acquisition, cleaning and transformation of data into useful information for the analyst. It should be highlighted that, before information can be effectively shown to the end user, the team in charge of data visualization must have a deep knowledge of the information they are transmitting, and its characteristics.

During the transition towards Web 2.0, there is a progressive increase in the huge amount of data that can be collected from it. Thus, the task of extracting this data is increasingly more complex. In order to solve this problem, there exists a number of data visualization tools, which will be detailed in later chapters of this Degree final project document. Some of these tools are Tableau Software and Google Analytics.

These tools are a fundamental weapon for analysts, who must know how to use it and determine whether the tool's ultimate goal is oriented to what the analyst wants to visualize or discover.

Thus, software companies have invested high amounts of money in the implementation of these tools, which are very often, oriented to the study of a particular subject or area. Because of this, it can be said that the task of the analyst also resides in being able to identify the ultimate goal of the visualization tool they are provided, in order to avoid that it turns into an obstacle which obstructs their analysis.

1.2. Motivations

Considering the previously explained context, starting at first with the rise of collaborative communities, we intend, at first, to provide the reader

with a theoretical background about the most successful collaborative projects, their characteristics, and existing studies and hypotheses related to them. In spite of focusing on explaining these kind of projects, the ultimate study focus of this degree final project is the not so successful, or even failed collaborative communities, whose study is not as widespread. In order to study this kind of communities, we aim to develop a tool that shows what happens in a collaborative community across time, and that it allows to understand how its underlying community of users and their participation evolves. It is also interesting to use this tool to extract a diagnosis of the state of the project, which can be used by the community itself, or by people interested in the wiki study area.

In order to create this tool, we aim to extend the existing tool of WikiChron Classic Serrano et al. (2018b) creating a new version of this tool, called WikiChron Monowiki. In this new version, we provide some metric definitions which allow the user to see aspects that remain hidden with the metrics already provided in WikiChron Classic. We will also improve the visualization modes that already exist in WikiChron Classic to help show these new aspects. This new version of the tool will allow it to observe deeper aspects of a single wiki, such as the composition of the active registered users community, or their activity in different wiki namespaces.

Moreover, as it was mentioned before, several existing studies are incomplete, as they focus in very popular projects, such as the ones carried out by Yang et al. (2016) or Kittur et al. (2006), which obtain conclusions that may not be applicable to other communities. Thus, it is interesting for us to use WikiChron Monowiki to show its utility and to apply some of those studies, or the one carried out by Serrano et al. (2018a) to Fandom communities.

Our code is available in GitHub¹, so other users can contribute and help to improve it in the future. They can also check any code they wish in case they have doubts.

1.3. Objectives

Following the previously explained context, and knowing the underlying motivations of our project, the objectives that drive it are stated below.

1.3.1. Study and research in collaborative communities area

At first, we intend to investigate existing studies and hypotheses, related to wiki repositories such as Fandom, like the one carried out in Serrano et al. (2018a), and to successful projects such as Wikipedia, like the ones carried out by Kittur et al. (2006) and Yang et al. (2016), in which they study how participation evolves in Wikipedia and Open Street Map respectively. The

¹<https://github.com/wikichron-tfg/WikiChron>

study carried out by Viégas et al. (2004) is also interesting, as they study the temporal evolution of pages, in terms of edits and content of some pages of the English Wikipedia.

1.3.2. Deciding important aspects and improve their visualization

We also aim to determine which aspects of a single Wiki are the most relevant, and which one is the best way to visualize them, in order to show the main information and omit irrelevant aspects. We will choose among four types of graphs: line graphs, heatmaps, bar graphs and line-area graphs. Each one of these graphs has their flaws and advantages, and we intend to find out which ones fit better to each aspect to visualize. Thus, we have extended WikiChron to provide different types of graphs.

1.3.3. WikiChron Monowiki

We aim to create a new version of an existing webtool, WikiChron Classic, that will be called WikiChron Monowiki. Both versions are willing to make it easier to generate a diagnosis of the state of a wiki in a particular time period, in order to use it with several objectives, such as to be used by the community, or to shed a light on the wiki study area. However, the main objective of WikiChron Monowiki is to show deeper aspects of a single wiki. We provide the code in a GitHub² repository, to make it easier to access it to future contributors.

1.4. Document structure

In the end, we provide a brief description of the chapters that compose the rest of the document:

- **Chapter 2** provides the reader with the theoretical framework needed not only to understand the history of collaborative communities, and particularly, the successful Wikipedia and the understudied Fandom, but also to know the existing studies and hypotheses concerning both collaborative projects.
- **Chapter 3** contains a set of existing tools to perform data visualization and analysis processes. Some of the tools exposed here are intended to analyze generic enterprise aspects. Some others are intended to analyze some wiki aspects. At last, in this chapter, we deeply describe WikiChron Classic, which is the starting point from which we began the creation of WikiChron Monowiki.

²<https://github.com/wikichron-tfg/WikiChron>

- **Chapter 4** provides a description of our contributions to WikiChron Classic.
- **Chapter 5** provides a description of WikiChron Monowiki and its main functionalities, including the motivations for its creation, the four new visualization modes that it includes.
- **Chapter 6** is a collection of Fandom wikis analyses, developed using WikiChron Monowiki. In this chapter, we intend to show the helpfulness of WikiChron Monowiki as an analysis tool, and also, to broaden existing research related to some Fandom wikis, and to analyze new aspects of these kind of communities.
- **Chapter 7** contains the conclusions of this work, along with the future work that can be done in order to improve the results obtained in this Degree final project.
- **Chapter 8** contains the work done by each one of the members of this project.

Capítulo 2

Marco teórico

2.1. El modelo de producción colaborativa

El modelo de producción colaborativa (en inglés, *commons based peer production*), es el área de estudio principal en nuestro trabajo de fin de grado. Este fenómeno, como se ha expuesto anteriormente, se basa en la cooperación libre y voluntaria, independiente del mercado, sin ánimo de lucro y sin bases establecidas alrededor de una remuneración económica, de acuerdo con Bates (2007). Los usuarios que deciden participar en este tipo de proyectos, lo hacen por voluntad propia, y sabiendo que su contribución es libre y al alcance de cualquier persona.

En su trabajo, Yonchai Benkler ha identificado dos características clave en la organización de este tipo de comunidades: modularidad y granularidad. La modularidad es una propiedad de un proyecto, por la que éste se puede dividir en pequeños módulos o componentes. Así, estos componentes se pueden producir de forma independiente, para ser posteriormente ensamblados en un “todo”. A su vez, la granularidad hace referencia al tamaño de un módulo, en términos del tiempo y esfuerzo que requiere su construcción. Estas dos propiedades favorecen la llegada a la comunidad de nuevos usuarios, siendo más fácil para ellos el contribuir a la creación de un módulo pequeño y de granularidad fina, para habituarse a las características del proyecto.

Estrechamente relacionado con estas dos propiedades de las comunidades que implementan un modelo de producción colaborativa, se encuentra el hecho de que la colaboración de sus participantes es geográficamente dispersa y está organizada en redes (Coffin, 2006). En concreto, se trata de proyectos que se basan en la idea de que las tareas y la elaboración e implementación de soluciones para diversos problemas o conflictos, se pueden llevar a cabo de forma descentralizada, desembocando así en mejores resultados (Benkler, 2017). Además, estas comunidades se basan en la idea de la equipotencialidad, es decir, que todos los individuos son, en el mismo nivel, potenciales

contribuidores en una comunidad (Bauwens, 2005) y los individuos tienen la capacidad de decidir en qué parte del proyecto contribuir.

Se trata de comunidades transparentes, en las que el diálogo entre los distintos usuarios se queda almacenado y es de libre acceso (e.g. Wikipedia tiene un historial de ediciones al que cualquier usuario tiene acceso). Esta transparencia, favorece el que estas comunidades se caractericen por seguir un método de organización por el que cualquier parte de la comunidad posee conocimiento horizontal sobre lo que está ocurriendo, y conocimiento vertical acerca de los objetivos del proyecto, difiriendo así del tradicional método de organización panóptico, en el que el conocimiento está distribuido jerárquicamente, de forma que tan sólo la cima de la pirámide tiene una visión al completo del proyecto (Bauwens, 2005). Con ello, cabe destacar que estas comunidades no poseen una estructura jerárquica de subordinación, sino que los usuarios se organizan para que su cooperación resulte más prolífica. No obstante, esto no significa que no haya ningún tipo de regulación dentro de estos proyectos, al contrario: en la mayoría de los casos, hay administradores y usuarios con privilegios que pueden gestionar los contenidos del proyecto (Siefkes, 2011).

Sin embargo, este método de organización estructural, que permite la existencia de varios grupos de colaboradores, participando en el proyecto en distintas direcciones, es considerado por algunos autores como un riesgo para estas comunidades, ya que pueden terminar convirtiéndose en jerarquías inflexibles. Así, esta forma de organización a menudo desemboca en una meritocracia en la que: “todo el mundo es libre para contribuir, pero las personas que gestionan el proyecto, son igualmente libres para rechazar esas contribuciones” (Karpf, 2009). En base a ello, Karpf (2009) afirma que existe un riesgo de que los contribuidores asumiendo este rol de gestores, abusen de su poder, lo que puede resultar en un éxodo de los miembros de la comunidad.

No obstante, existen ejemplos de comunidades cuya estructura interna ha evolucionado hasta una “dictadura benevolente”: proyectos de software libre como Linux, en el que el rol de gestor supremo es asumido por Linus Torvald, o proyectos como Wikipedia, en el que este rol lo asume Jimmy Wales. No obstante, estudios como Shaw y M. Hill (2014) demuestran que este rol puede ser asumido no sólo por una única persona, sino por una pequeña élite. Algunos autores como Coffin (2006), consideran la existencia de estos roles como algo positivo y necesario, puesto que en el modelo de producción colaborativa es importante tanto la participación como el saber gestionar las colaboraciones de manera constructiva para el proyecto.

Respecto a las comunidades colaborativas, existen estudios que generalizan el comportamiento de los contribuidores en cuanto a su participación, para la mayoría de ellos. Uno de estos estudios es la regla del 90-9-1, Nielsen (2006):



Figura 2.1: Regla del 90-9-1.

La regla del 90-9-1 según Nielsen (2006), se basa en el principio de que, en la mayoría de las comunidades colaborativas, una pequeña minoría de usuarios asume la mayor parte del trabajo. Así, en líneas generales, el comportamiento de los usuarios de la comunidad, sigue una regla de 90-9-1 (ver figura 2.1), en la que:

- El 90 % de los usuarios tan sólo se lucran con el contenido generado por la comunidad, o participan de manera indirecta, como por ejemplo difundiendo la existencia de la comunidad y aumentando así, su visibilidad.
- El 9 % de los usuarios contribuyen de vez en cuando, pero durante la mayor parte del tiempo que pasan en la comunidad, sus prioridades son otras.
- El 1 % de los usuarios generan una cantidad descomunal de trabajo, asumiendo, así, la creación de la mayor parte del contenido existente en la comunidad.

La mayoría de las comunidades colaborativas siguen un patrón por el estilo, o incluso más abrupto. Por ejemplo, la regla seguida por Wikipedia es un 99,8-0,2-0,03, de acuerdo con Nielsen (2006).

Así, ante la pregunta de cómo evitar la desigualdad en la participación de las comunidades colaborativas, Nielsen responde que la desigualdad en contribuciones es algo que no se puede evitar del todo, tan sólo, suavizar por medio de acciones como el recompensar, en su justa medida, a los usuarios que sí contribuyen, favorecer un mecanismo que permita contribuir de forma sencilla, y promover las contribuciones de calidad, a pesar de que no pertenezcan a usuarios que se encuentran dentro del 1 % de contribuidores más activos.

De esta manera, se puede conseguir una participación menos desequilibrada dentro de las comunidades colaborativas, y evitar problemas como la falta de representatividad del 90 % de los usuarios que no contribuyen en

elementos de la web como motores de búsqueda o sistemas de recomendación basados en rankings, en los que la mayoría de los votantes pertenecen al 1 % de usuarios más activos.

2.2. La tecnología Wiki

Leuf y Cunningham (2001), creadores de la primera wiki, definen una wiki como una colección de páginas web entrelazadas, libremente expandible, y un sistema de hipertexto para almacenar y modificar información. En concreto, se trata de una base de datos en la que cada página puede ser editada fácilmente por cualquier usuario con un buscador web adecuado. Además, estos autores establecen una diferencia entre el término Wiki (con mayúscula) y wiki:

Wiki hace referencia a este tipo de tecnología, mientras que wiki, hace referencia a una aplicación en particular. Así, las wikis son sitios web interactivos, que le otorgan al usuario la capacidad para modificar o crear el contenido de una página, o borrar dicho contenido, sin que el usuario en cuestión tenga ningún tipo de conocimientos o habilidades previas en la publicación de contenidos en la web. Se introducen, además, las diferentes funcionalidades para las que puede ser utilizada una Wiki (Raman, 2006):

- Una Wiki se puede utilizar como un sistema para administrar contenido.
- Además, se puede utilizar como foro de discusión, o para cualquier otro tipo de trabajo en grupo.

La tecnología Wiki se basa en software libre. Así, el software en el que se basa una wiki es llamado un motor de wiki. Existen diversos motores de wiki libres, llamados clones de wiki, o “wiki clone” (Leuf y Cunningham, 2001). Algunos de estos ejemplos son los siguientes:

- Wiki: es la wiki original creada por Ward Cunningham, con diversos recursos disponibles en la web, y que es sencilla de instalar.
- Twiki: diseñada para ser empleada por usuarios corporativos. Se define como más profesional y poderosa, con funcionalidades como e-mails automáticos, y un sistema que permite adjuntar ficheros.
- Plone: implementa un sistema de almacenamiento de ficheros, proporcionando así, una implementación que apoya una gran demanda corporativa. Además, proporciona una funcionalidad por la que permite descargar PDFs online de manera sencilla y simple.
- TikiWiki: Se trata de una wiki sencilla de instalar y usar, similar a la inicialmente creada por Cunningham.

- MediaWiki: es un motor de wiki empleado por todas las wikis englobadas por WikiMedia, entre ellas, Fandom, objeto de estudio de este trabajo, y Wikipedia. Este software permite modificar y difundir conocimiento, además de ser personalizable para las necesidades específicas de cada wiki.

Originalmente, la tecnología Wiki no estaba orientada a comprometer y captar la atención de usuarios casuales, sino que pretendía permitir la edición de páginas regularmente a los usuarios, siguiendo una dinámica colaborativa. El autor Wagner (2004), identifica los once principios fundamentales de diseño que caracterizan el diseño funcional de una wiki, entre los que destacan la transparencia y su libre acceso.

Finalmente, la tecnología Wiki es caracterizada por Gupta y Sushil (2004) como un sistema de manejo de conocimiento de tipo “groupware”. Los sistemas basados en la tecnología groupware, son sistemas basados en el uso de un ordenador para favorecer la creación, difusión y modificación de conocimiento a través de las tecnologías de la información. Otro tipo de sistemas basados en esta tecnología son los blogs, sistemas de mensajería instantánea, o los sistemas de videoconferencia.

2.3. Wikipedia

A continuación, se expone el marco teórico de uno de los proyectos colaborativos más populares en la actualidad. Es interesante conocer las características de Wikipedia, y de algunos de los estudios que existen sobre ella, con el objetivo de familiarizarnos mejor con la dinámica de las comunidades colaborativas existentes. Más adelante en este capítulo, estudiaremos las características de Fandom, para conocer también la dinámica de este otro tipo de comunidades colaborativas.

Wikipedia surge en el año 2001 fundada por Jimmy Wales, como un experimento paralelo a su predecesora, Nutpedia, que se basaba en la escritura enciclopédica, es decir, en la publicación de artículos escritos por expertos, tan sólo tras un largo período de revisión y edición. El primer objetivo de Wikipedia fue el permitir la colaboración libre en la creación de artículos, antes de que estos entraran en este proceso de revisión (Viégas et al., 2004). Poco a poco, este enfoque fue atrayendo a cada vez más contribuidores, hasta que se fueron creando las distintas Wikipedias en varios idiomas, culminando este proceso en Marzo de 2005, momento en el que se alcanzaron los 195 idiomas. Esta enciclopedia on-line, reúne más de 15.000.000 artículos, y se encuentra dentro del top 10 de sitios web más visitados (Voss, 2005).

Esta comunidad colaborativa exhibe, además, la característica de transparencia típica de este tipo de proyectos: su historial completo de ediciones es accesible para todo el público. Por este hecho, además de la gran canti-

dad de datos presentes y el éxito de la comunidad, existen múltiples debates acerca de diversos factores de Wikipedia.

2.3.1. Estructura de la comunidad y gobernanza

El primer factor a tratar sobre la gobernanza en Wikipedia, es que algunos autores debaten sobre la existencia, en Wikipedia, de lo que llaman “una dictadura benevolente”, cuyo líder es Jimmy Wales, su creador (Kostakis, 2010). Autores como Coffin (2006), defienden la naturaleza positiva de la misma, en base a la necesidad de mantener un orden dentro de todo proyecto colaborativo: no sólo es esencial para la supervivencia de una comunidad colaborativa el atraer a usuarios que colaboren en ella, sino también, el garantizar que estas colaboraciones sean óptimas para la proliferación del proyecto, dado que la facilidad para contribuir en él, atrae ocasionalmente a usuarios con intención de dañarlo. A su vez, Pink (2005), coincide con Coffin (2006) en que el rol originalmente asumido por Jimmy Wales era una medida necesaria para proteger el proyecto de actos vandálicos que podían dañarlo. Así, surgió en Wikipedia una jerarquía de usuarios en función de los permisos que éstos tienen, y de las acciones que puedan llevar a cabo. Esta red de permisos se describe como una pirámide por Pink (2005), cuyos niveles están conformados por los diferentes tipos de usuarios:

- En la base, se encuentran los usuarios anónimos, identificados única por una dirección IP, y que, según Pink, llevan a cabo unas pocas ediciones, Pink (2005).
- En el siguiente nivel, se encuentra un gran número de usuarios registrados de Wikipedia, que llevan en la comunidad cierto tiempo, y han alcanzado cierto número de ediciones. Una vez alcanzados estos requisitos, se les conceden permisos para crear nuevos artículos, mover páginas, subir archivos y editar páginas protegidas o semi-protegidas, según Wikipedia user access levels.
- Posteriormente, y tras un período de alta dedicación a la comunidad como usuarios registrados comunes, se encuentran los administradores. Este tipo de usuarios tiene permisos para borrar artículos, proteger páginas, editar páginas protegidas y bloquear y desbloquear direcciones IP. Para convertirse en administradores, tienen que pasar en primer lugar por un proceso de selección, al que entran o bien por voluntad propia, solicitándolo, o bien porque son nominados por otros usuarios. Este proceso es conocido en Wikipedia como “solicitud de administración”, cuya duración es de siete días desde el momento en el que se presenta la solicitud. Durante ese tiempo, se lleva a cabo, un estrecho seguimiento de su actividad como editores, con el objetivo de descartar las solicitudes fraudulentas (e.g usuarios que son nuevos en Wikipedia),

para posteriormente, llevar a cabo un proceso en el que se le formulan diversas preguntas al candidato, a través de la página de su Solicitud de administración, que éste ha de responder. Finalmente, tras este período de tiempo, los burócratas (usuarios que se encuentran en un nivel superior al de los administradores), examinan esta página, y le asignan un porcentaje a la Solicitud de administración del candidato. Si este supera un 65, se le asigna el rol de administrador (Wikipedia request for adminship).

- En el siguiente nivel, se encuentran los usuarios burócratas, grupo que está dotado de permisos para llevar a cabo determinadas acciones sobre las cuentas de otros usuarios. En concreto, los burócratas tienen la capacidad de eliminar a usuarios de la categoría de administradores. También tienen permisos para hacer a otros usuarios administradores, y burócratas. No obstante, nótese que los burócratas no tienen la capacidad de eliminar a otros usuarios de la categoría de burócratas. Para entrar en esta categoría, de forma similar al proceso para ser administrador, existe en Wikipedia lo que se conoce como Solicitud de burócrata (en inglés, Request for Bureaucratship) (Wikipedia user access levels).
- Por encima de los burócratas comunes, se encuentra un grupo reducido de burócratas altamente privilegiados, los desarrolladores, que pueden llevar a cabo cambios en el software y la base de datos de Wikipedia. A diferencia de los grupos anteriormente descritos, este tipo de usuarios forma parte del equipo de Wikimedia Foundation (Wikipedia developers).
- Por último, en lo alto de la pirámide, al principio se encontraba el grupo de fundadores, al que sólo pertenecía Jimmy Wales, que tenía todos los poderes persistentes, desde crear cuentas, bloquear a usuarios y acceder a revisiones eliminadas y eliminar a usuarios de todas las categorías. No obstante, cualquier acción llevada a cabo ha de ser previamente consensuada. En Mayo de 2010, los permisos de Jimmy Wales fueron reducidos porque eliminó algunos archivos de Wikimedia sin haberlo consensuado primero (Wikipedia role of Jimmy Wales), de manera que en la actualidad, hay un comité de la WikiMedia Foundation, que establece un equipo directivo para cada una de las Wikipedias.

En segundo lugar, existen, en Wikipedia, algunos conflictos que se relacionan intrínsecamente con la característica inherente de Wikipedia: el hecho de que cualquier contribuidor que así lo desee, puede participar, y con esta red de permisos. Este, en concreto, es un conflicto que surge de la mano de los administradores de Wikipedia, Kostakis (2010), y se trata de una disputa acerca del destino de Wikipedia: Por una parte, se encuentran los llamados

“inclusionistas”, que apoyándose en las bases, también defendidas por los autores Viégas et al. (2004), de que Wikipedia no es una enciclopedia en papel, por lo que no es estática ni está espacialmente delimitada, defienden una Wikipedia que debe acumular tanto conocimiento como le sea posible. Por otra parte, se encuentran los “deleccionistas” (del inglés, “Delectionists”), que defienden una Wikipedia más exclusiva, sin artículos sobre temas triviales y en la que la información existente sea revisada, contrastada y mejorada, *battle for Wikipedia’s soul* (2008). Esta segunda ideología, es también compartida por Larry Sangers, cofundador de Wikipedia, quién, en Sangers (2004), describe dos grandes problemas de Wikipedia:

El primero, es que Wikipedia no se percibe como una fuente fiable de conocimiento, dado que cualquier contribuidor tiene acceso a ella y puede añadir una entrada con desinformación. Así, Wikipedia publicará dicha entrada sin llevar a cabo ningún tipo de revisión oficial para verificar la información expuesta en ella.

El segundo problema de Wikipedia, según Sangers (2004), es su “anti elitismo”. Según su punto de vista, es contraproducente para Wikipedia el que investigadores con un gran conocimiento sean colocados al mismo nivel que cualquier usuario con acceso a la web.

Así, dadas estas dos corrientes de pensamiento entre usuarios de Wikipedia, se originan multitud de disputas sobre si distintos artículos deben ser admitidos o no en Wikipedia. Según un artículo del periódico *The New York Times*, Schott (2009), la presencia de los “Deleccionistas” habría provocado, en 2009, un crecimiento más lento de Wikipedia.

La jerarquía de derechos de usuario anteriormente expuesta, y el surgimiento de disputas como la de “deleccionistas vs inclusionistas”, generan críticas hacia Wikipedia, tildándola de ser una “tiranía”, en la que es extremadamente difícil lidiar con los usuarios más privilegiados que no se adhieren al contrato social de la comunidad, Kostakis (2010).

2.3.2. Conflictos en la comunidad

En sus inicios, algunos consideraron la facilidad de contribución en Wikipedia, que permite que cualquier usuario con acceso a la web pueda manipular o crear un artículo, como una desventaja, llevándolos a poner en duda el que Wikipedia se convirtiera en una herramienta de referencia. Así, para algunos investigadores, el misterio de Wikipedia reside en cómo, a pesar de esta desventaja, la comunidad ha logrado un éxito tan rotundo.

Esta propiedad promueve el que Wikipedia se encuentre en continuo cambio, lo que puede ser una fuente de fortaleza, ya que no se trata de una enciclopedia estática, sino que su dinamismo le permite recoger conocimiento sobre las áreas más populares de nuestra sociedad en cada momento. Pero también, puede ser una fuente de debilidad, dado que cualquier usuario pue-

de contribuir y dañar el proyecto, Viégas et al. (2004). De esta manera, surge uno de los problemas más frecuentes en Wikipedia: el vandalismo.

Según la literatura, el número de usuarios con intenciones dañinas para la comunidad, se multiplica a medida que el proyecto crece. Así, surge la pregunta: ¿cómo puede ser que todavía no hayan saturado el proyecto con contribuciones dañinas? La respuesta la dan los autores Viégas et al. (2004), haciendo uso de dos funcionalidades clave de Wikipedia:

En primer lugar, la comunidad ofrece, para los usuarios registrados, las llamadas “Watch lists” o listas de vigilancia. En ellas, el usuario puede añadir todas las páginas en las que esté interesado, para observar qué cambios ocurren en ellas. Las listas de vigilancia, ofrecen, por página añadida, una lista de los cambios más recientes hechos en la página. Además, el usuario recibirá un email cada vez que una de las páginas incluidas en su lista de vigilancia cambie (Wikipedia Watchlist). Esta funcionalidad permite que un usuario pueda revertir un acto vandálico sobre cualquiera de las páginas en su lista de vigilancia de forma rápida.

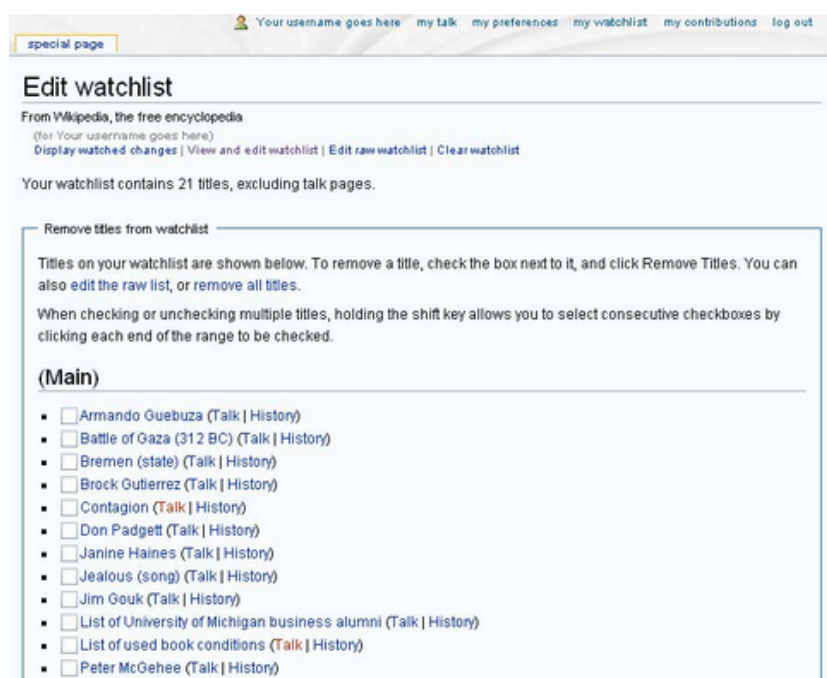


Figura 2.2: Lista de vigilancia de Wikipedia

Por otra parte, el conjunto de datos que almacena el histórico de Wikipedia, contiene, para cada página, todas las versiones que se han hecho de la misma, lo que incluye el texto de la propia página en cada versión (Wikipedia database download).

Esta investigación utiliza el conjunto de datos de Wikipedia, para estu-

diar, en primer lugar, el tiempo medio que tarda Wikipedia en deshacer actos vandálicos. Las denominadas “ediciones maliciosas”, Viégas et al. (2004), son extremadamente frecuentes y variadas en proyectos colaborativos de éxito como lo es Wikipedia. No obstante, los autores Viégas et al. (2004), han identificado los cinco tipos de ediciones maliciosas más comunes en Wikipedia:

- Borrado en masa: se trata de la eliminación de todos los contenidos de una página.
- Copia ofensiva: inclusión de términos vulgares o calumnias, en los contenidos de una página.
- Copia falsa: inclusión de contenido que no está relacionado con el tema tratado en la página en cuestión.
- Redirección falsa: inclusión de links que redireccionan al lector a contenido que no está relacionado con el tema tratado en la página, o que es ofensivo.
- Copia idiosincrásica: inclusión de contenido no imparcial, o que no es de interés general. Un ejemplo es añadir un texto de gran extensión exponiendo oraciones religiosas del Corán en la página del Islam.

En su investigación, Viégas et al. (2004) descubren algunos fenómenos interesantes, como que la mitad de los llamados borrados en masa, son corregidos en menos de 2 minutos, o que la mitad de los borrados en masa que además contenían palabras obscenas, en menos de 1,5 minutos. Así, se puede concluir que Wikipedia, a pesar de ser muy vulnerable a las ediciones maliciosas, debido a la facilidad que otorga a cualquier usuario para contribuir, tiene también una gran capacidad de recuperación, gracias a las llamadas listas de vigilancia y a la acción de los usuarios realmente comprometidos con el proyecto.

En segundo lugar, Viégas et al. (2004) investigan, empleando el histórico de versiones de diversas páginas, otro fenómeno: las disputas de ediciones. En Wikipedia, existen disputas entre los contribuidores, no sólo en temas controvertidos, sino en todo tipo de páginas. Una disputa puede ser del calibre de si un párrafo debe ser incluido o no en una página, dando lugar a varias versiones de la página, que quedan almacenadas en el histórico de Wikipedia. El descubrimiento de estos autores reside en que los usuarios emplean los comentarios adicionales que se pueden añadir explicando los cambios hechos en la edición de una página, para mantener conversaciones, que incluyen preguntas, y frases especialmente dirigidas de un contribuidor a otro, como si se tratara de un foro. En su estudio, formulan la hipótesis de que los usuarios tratan de resolver las disputas a través de estas descripciones en las ediciones de la página. No obstante, cuando no consiguen convencer a los demás de su punto de vista, la discusión trasciende a las páginas de debate de Wikipedia.

2.3.3. Participación

Una de las preguntas más formuladas acerca de Wikipedia es de dónde viene su éxito, es decir: ¿es este éxito un resultado del conocimiento de las masas? ¿O es, por el contrario, fruto del poder de unos pocos usuarios que contribuyen en grandes cantidades de ediciones, como se plantea en Kittur et al. (2006)?

La respuesta a esta pregunta es considerada de gran utilidad para la propia Wikipedia, y para otros sistemas colaborativos nuevos, similares a ella. Según Kittur et al. (2006), a la hora de tomar las decisiones de diseño para construir una plataforma para una comunidad colaborativa, es importante tener en cuenta a aquellos grupos de usuarios que son más influyentes. Así, empleando el histórico de Wikipedia hasta Enero de 2006, estos autores se decidieron a proporcionar una respuesta. En primer lugar, realizaron un estudio sobre la actividad de los administradores a lo largo del tiempo, cuyos resultados son que el número de contribuciones realizadas por administradores dividido entre el número de contribuciones totales realizadas en un mes, comienza a decrecer a comienzos del año 2004. Para descubrir a qué se debe esta decaída, formulan las siguientes hipótesis:

- En primer lugar, puede deberse a que haya una decaída en el número de ediciones realizadas por administradores.
- En segundo lugar, puede deberse a que la carga de trabajo de los administradores decrezca gracias a la acción de los bots de Wikipedia, creados para lidiar con actos vandálicos.
- En tercer lugar, es posible que esta decaída se deba a un aumento en las contribuciones realizadas por el resto de usuarios de Wikipedia, que no son administradores.

Así, para determinar cuál de las tres hipótesis es verdadera, llevan a cabo tres estudios:

- En primer lugar, observan que el número de ediciones realizadas por administradores se encuentra en continuo crecimiento, por lo que la primera hipótesis queda definitivamente refutada.
- En segundo lugar, la actividad de los bots sigue una línea prácticamente constante, con pequeñas variaciones, que no puede explicar la decaída en el número relativo de contribuciones de los administradores.
- Por último, crean cinco categorías de usuarios en base a su número de ediciones, con el objetivo de clasificar a los usuarios no en base al estatus otorgado por Wikipedia, que puede variar con el tiempo, sino en base a una medida robusta. En concreto, las clases son las siguientes: usuarios con menos de 100 ediciones, usuarios con un número de

ediciones mayor o igual que 100 y menor que 1000, usuarios con un número de ediciones de entre 1000 y 5000, usuarios con un número de ediciones de entre 5000 y 10.000, y por último, usuarios con más de 10.000 ediciones hasta el momento. Así, estudian la productividad en número de contribuciones al mes de cada una de estas categorías de usuarios, llegando a la conclusión de que la clase de usuarios con menos ediciones (en el caso de este estudio, ≤ 100 ediciones), es la que eclipsa con su trabajo a la clase de usuarios con más ediciones, siendo así, la responsable de esta decaída.

Por tanto, con este estudio, los autores concluyen que, en 2004, comienza en Wikipedia un efecto al que llaman “la sabiduría de las masas”, por el que gran parte del trabajo de Wikipedia es realizado por usuarios que no pertenecen a la élite. Este artículo evidencia la utilidad de categorizar a los usuarios en función del número de ediciones que han hecho en la wiki, estrategia que seguimos en WikiChron Monowiki, ya que tal y como defienden estos autores, es más útil categorizar a los usuarios en función de medidas disponibles o que se pueden calcular a partir del histórico de la wiki, que categorizarles atendiendo a criterios hechos por la propia comunidad, ya que son subjetivos y pueden variar con el tiempo. En este trabajo, hemos hecho un análisis con dos wikis de Fandom que sigue una línea similar a la de Kittur et al. (2006), desgranando qué grupos de usuarios son los más contribuyentes en ellas.

2.3.4. Evolución del contenido

A parte del nivel de participación de los diferentes tipos de usuarios a lo largo del tiempo, existen estudios que tratan de establecer qué eventos pueden influenciar el nivel de participación de los grupos de usuarios más activos y cómo afecta esta participación al contenido de los artículos (Chhabra y Iyengar, 2018; Lin y Chen, 2018). Por otra parte, existen estudios acerca de cómo afectan determinados eventos a la participación de los grupos de usuarios menos activos, como Fangzhou Zhang et al. (2019).

En Lin y Chen (2018), se justifica, en base a los datos de la Wikipedia China, que el nivel de participación de los usuarios más activos, en ocasiones es provocado por las contribuciones de los usuarios más nuevos y menos activos. Estas intervenciones son causadas especialmente por un aumento en la frecuencia de la participación de los usuarios menos activos en un artículo. Además, estas ediciones resultan en una mayor calidad para el artículo en cuestión.

Siguiendo estos estudios, en Chhabra y Iyengar (2018), se especifica que todo artículo comienza compuesto por una serie de factoides, que posteriormente, son leídos por otros contribuidores, desencadenando así el que éstos añadan más información, y así sucesivamente. De esta manera, en Wikipedia, el contenido existente desencadena más participación por parte de otros

usuarios, dando lugar a la evolución continua de los artículos por medio de ediciones sucesivas. Estos hechos tienen una explicación psicológica, basada en las perturbaciones o conflictos cognitivos de la mente del ser humano, que se activan cuando este ve contenido que contiene información que, en su mente, es más completa o tiene un enfoque distinto. En dicha situación, el individuo en cuestión sentiría una necesidad (más o menos intensa en función de diversos factores) de contribuir a completar dicho conocimiento.

La investigación llevada a cabo en este artículo, se basa en que el conocimiento está compuesto por nodos de información, más pequeños, interconectados unos con otros por medio de aristas, cuyo grosor determina la fuerza de asociación entre dos nodos de información o conceptos.

Con el objetivo de comprender cómo funciona la evolución del conocimiento en Wikipedia, los autores del último artículo diseñan una forma de detectar los factoides en un artículo de Wikipedia:

Los factoides son piezas de información básica para describir la idea que se está tratando. A su vez, si un artículo existe en Wikipedia, es porque tiene cierta relevancia. En base a estas dos últimas ideas, los investigadores anteriores, establecen que, en un artículo:

- Los links que redirigen a otros artículos son factoides.
- También son factoides aquellos términos que sobrevivan hasta las últimas revisiones del artículo.

Para computar el grosor de las aristas que unen estos conceptos, en Chhabra y Iyengar (2018), se emplea la Distancia de Google normalizada, que es una forma de computar la similitud semántica entre dos conceptos. Esta técnica se basa la idea de que dos frases semánticamente relacionadas, se encontrarán juntas en un mayor número de resultados en Google.

Empleando esta técnica, los autores reúnen, por cada artículo estudiado, una lista de listas, en la que cada elemento es una lista de la forma:

[revisión(i), factoides(i), revisión(j), factoides(j)]

En el que revisión(i) y revisión(j) son revisiones llevadas a cabo por distintos usuarios, y factoides(i) y factoides(j) son los factoides añadidos entre una revisión y otra. Así, analizan el valor de la Distancia de Google Normalizada para los factoides existentes entre revisiones sucesivas de un artículo, efectuadas por contribuidores distintos. Para encontrar qué factoides están muy relacionados, computan el producto cartesiano entre los factoides(i) y los factoides(j), y computan el valor de la Distancia de Google Normalizada para todos los pares. Aquellos pares de factoides con un valor de NGD bajo, son los que mayor relación semántica tienen. Así, los resultados que obtuvieron para dos artículos de Wikipedia, evidencian que la inclusión de determinados términos en una revisión, está estrechamente relacionada con la inclusión de otros términos determinados en una revisión posterior.

Continuando con el estudio de fenómenos que influyen la participación en artículos existentes de Wikipedia, Fangzhou Zhang et al. (2019) habla de otro fenómeno que desencadena la participación de usuarios en un artículo: los llamados shocks de atención, fenómenos que ocurren en el mundo real, y que pueden atraer una participación repentina sobre un artículo existente determinado, relacionado con dicho fenómeno.

Estos “shocks”, son descritos como un factor de gran importancia para conseguir el compromiso de nuevos usuarios para con un artículo determinado, pero también pueden repeler a estos usuarios. Así, Fangzhou Zhang et al. (2019) estudia cuáles son los factores que hacen que usuarios no tan activos en Wikipedia sean retenidos por un artículo particular en el largo plazo. (con no tan activos, se refieren a usuarios que ya eran contribuidores de Wikipedia, aunque su participación no es notable).

En este artículo se llevan a cabo las siguientes hipótesis en cuanto a la la participación, cuando un shock de atención impacta a un artículo:

- En primer lugar, se formula la hipótesis de si el impacto de un shock de atención sobre un artículo, está relacionado con un aumento en el feedback negativo de la comunidad hacia los contribuidores nuevos del artículo.
- Posteriormente, se formula la hipótesis de si el impacto de este tipo de fenómeno, está relacionado con la centralización en las contribuciones de un artículo (la idea de que cada artículo es elaborado por un pequeño porcentaje de editores).
- Más tarde, se formula la hipótesis de si el impacto generado por un shock de atención sobre un artículo, está relacionado con un aumento en las páginas de discusión de dicho artículo.
- Además, se debate sobre si el feedback negativo de la comunidad hacia los usuarios nuevos, está relacionado positiva o negativamente con el nivel de participación.
- Por otra parte, se debate sobre si la centralización está positivamente relacionada con la participación de nuevos usuarios en un artículo.
- Siguiendo con el debate, se especula sobre si la discusión está positivamente relacionada con la participación, es decir, si el observar las páginas de discusión es una forma positiva de motivar a los usuarios nuevos y enseñarles las normas de la comunidad.
- Finalmente, se debate también sobre si la participación de los usuarios nuevos del artículo en las páginas de discusión está positivamente relacionada con su participación a posteriori.

De acuerdo con los resultados obtenidos por Fangzhou Zhang et al. (2019), los shocks de atención tienden a atraer a un gran número de participantes a un artículo. Tras analizar la actividad de estos nuevos participantes, observan que estos tienen menor experiencia en Wikipedia que aquellos contribuidores que ya se encontraban editando el artículo de forma comprometida antes del shock de atención. Así, proponen la posibilidad de que al ser atraídos al artículo por el shock de atención, estos usuarios tienen la oportunidad de interactuar con los usuarios que ya estaban comprometidos con el artículo antes del shock y aprender de ellos, ya que los usuarios ya comprometidos, por lo general, según sus observaciones, tienen más experiencia en Wikipedia.

Por otra parte, observan que el nivel de centralización en el artículo, decrece durante el shock de atención, lo cual tiene sentido, ya que hay más trabajo a realizar, y más contribuidores dispuestos a asumir la carga de tareas. En cuanto al feedback negativo en forma de borrado de ediciones, se observa que, en general, los shocks de atención son un período caótico para el artículo, por lo que el número de borrados aumenta. No obstante, dado que los usuarios nuevos en el artículo, de acuerdo con las observaciones previas, suelen tener menos experiencia en Wikipedia, son un núcleo mayor de borrado de ediciones que los usuarios que ya estaban comprometidos con el artículo.

Por otra parte, se observa que durante el shock de atención, hay una mayor edición de las páginas de discusión, dado que es necesario coordinarse en un momento en el que hay más cantidad de usuarios y de trabajo a llevar a cabo. No obstante, no hay un crecimiento notable en el nivel de contribuciones de los usuarios nuevos en el artículo a la página de discusión. Tras llevar a cabo un estudio estadístico de estas observaciones, se concluye lo siguiente:

- En primer lugar, se confirma que los shocks de atención atraen a una gran cantidad de participación, que es necesario gestionar y coordinar.
- En segundo lugar, el feedback negativo está relacionado con una mayor participación futura de los usuarios nuevos en el artículo. Se sugiere que este hecho puede estar relacionado con que estos usuarios nuevos, ven el feedback como un camino de aprendizaje y mejora.
- Por último, se observa que un mayor número de contribuciones en la página de discusión, está positivamente relacionado con una mayor participación futura de los usuarios nuevos en el artículo. Los autores sugieren que las páginas de discusión son una buena herramienta para que estos nuevos usuarios socialicen y adquieran un “pensamiento comunitario”.

2.4. Fandom

A continuación, se expone el marco teórico de la comunidad de Fandom, o Wikia. El conocimiento de las características de este proyecto colaborativo, junto con los estudios ya existentes sobre el mismo, son relevantes para este trabajo de fin de grado, en primer lugar, porque conocer el producto a analizar es de suma importancia a la hora de asegurar que las funcionalidades de WikiChron Monowiki cubren las necesidades del usuario final, el analista de wikis. En segundo lugar, porque en este trabajo de fin de grado, es de interés el emplear WikiChron Monowiki para continuar estudios ya existentes sobre Fandom, y dar lugar al inicio de nuevas hipótesis acerca de su estructura y funcionamiento.

Fandom, conocido antes de 2016 como Wikia, es un servicio de alojamiento de páginas web, todas ellas basadas en la tecnología Wiki. Al igual que Wikipedia, está basada en el software de MediaWiki.

Wikia fue fundada en 2004 por Jimmy Wales y Angela Beesley, ambos miembros de la fundación Wikimedia, *esacademic.com* (2019), y en la actualidad, aloja a más de 385.000 wikis, y en total, tiene más de 50.000.000 páginas (Fandom explore).

Un 46.9 % de la audiencia de *wikia.com*, proviene de Estados Unidos, un 6.8 % desde China, y un 3.3 % desde Canadá. Además, un 31 % de los visitantes de *wikia.com* están interesados en la animación, siendo las wikis más prolíficas en este aspecto: *myanimelist.net* y *crunchyroll.com*, (*Alexa site information*). A su vez, un 86 % de la audiencia está interesada en videojuegos, siendo las wikis más prolíficas las siguientes: *gamepedia.com*, *gamespot.com* y *steamcommunity.com*. Fandom es definida por la propia comunidad como “la voz de los fans del entretenimiento”, y “un lugar en el que los fans pueden discutir, encontrar y compartir conocimientos sobre cualquier tema” (Fandom explore).

No obstante, su fin no es únicamente la creación y difusión de conocimiento, sino que pretende, además, proporcionar un medio de interacción para los aficionados de una misma temática.

2.4.1. Estructura de la comunidad

A continuación se proporciona un trasfondo teórico sobre las diferentes categorías de usuarios que componen la comunidad de Fandom:

- En primer lugar, se encuentran los usuarios anónimos que, al igual que en Wikipedia, pueden editar y modificar el contenido de una página. Este tipo de usuarios es identificado por su dirección IP. No obstante, Fandom proporciona una funcionalidad para deshabilitar la edición de contenido por parte de usuarios anónimos.
- En segundo lugar, se encuentran los usuarios registrados, que pueden

personalizar la apariencia de la comunidad, subir una imagen, vídeo u otro tipo de archivo. Además, al igual que en Wikipedia, los usuarios registrados pueden añadir páginas a su lista de vigilancia, para observar los cambios en páginas que son de interés para el usuario. Además, pueden tener y personalizar su perfil de usuario, y eliminar los anuncios de publicidad de todas las páginas, excepto de la página principal de la wiki.

- En tercer lugar, Fandom distingue a otro tipo de usuarios, los usuarios autoconfirmados. Estos usuarios son aquellos que lleven registrados en Fandom durante, al menos, cuatro días. Tienen acceso a herramientas como el poder editar páginas semiprotegidas, y evitar pasar por un captcha al insertar un link externo o eliminar todos los contenidos de una página.
- Posteriormente, se encuentran los moderadores de contenido, que tienen acceso a algunas herramientas que les permiten modificar determinadas partes de la wiki. Estas herramientas incluyen eliminar y mover páginas protegidas, eliminar y mover archivos, revertir el borrado o ediciones en páginas y archivos, proteger y desproteger páginas, observar si una nueva edición en una página ha sido revisada o no. Esta última acción es posible si la comunidad tienen activa la funcionalidad “RecentChanges patrol”.
- A continuación, se encuentran los moderadores de discusión, que poseen herramientas para moderar la interacción entre usuarios en distintos lugares de la wiki. Entre estas herramientas, se encuentran: eliminar y restaurar hilos de conversaciones entre usuarios, mover estos hilos de conversación de un lugar a otro, eliminar comentarios en el blog de la wiki, y moderar el foro de discusión de la wiki, en caso de que esta tenga dicha funcionalidad. El foro de discusión de la wiki es un lugar especialmente habilitado para mantener conversaciones con otros usuarios sobre la temática tratada en la comunidad.
- Por otra parte, existe otro tipo especial de usuarios, los moderadores de chat: estos usuarios poseen herramientas que les permiten moderar la “chat room” de la comunidad. Entre estas herramientas, se encuentra el prohibir a determinados usuarios al entrada y participación en el chat.
- A continuación, se encuentran los administradores, o System Operators (sysops), que son usuarios de confianza para la comunidad, y que son elegidos de manera democrática. Tienen derecho a llevar a cabo acciones como bloquear a usuarios que están llevando a cabo actos de vandalismo para que no editen la wiki, eliminar a usuarios de la categoría de moderador de chat o moderador de discusión, editar el

formato y apariencia de la comunidad. Además, estar en la categoría de administrador implica, automáticamente, poseer los derechos de las categorías de moderador de discusión y moderador de contenido.

- Posteriormente, se encuentran los burócratas, que tienen el derecho a manipular los derechos del resto de clases de usuarios. Además, pueden bloquear y desbloquear a usuarios, al igual que los administradores. Por otra parte, pueden eliminar a usuarios de las categorías de Rollback, moderador de contenido y administrador. Así mismo, pueden crear nuevos burócratas. No obstante, no pueden eliminar a otros usuarios de la categoría de burócratas. Además, pueden eliminar a usuarios de la categoría de bots.
- Por último, se encuentran los fundadores, que son los creadores de la wiki, y asumen, automáticamente, derechos de administrador y de burócrata.

En Fandom, hay más tipos de usuarios además de los anteriormente citados. Entre ellos se encuentran los bots, que son procesos automáticos que le permiten a un usuario el llevar a cabo ediciones o acciones repetitivas en la comunidad. Así, si un usuario desea crear un bot, debe comunicarlo en la comunidad y obtener aprobación. Una vez obtenida la aprobación, el usuario en cuestión deberá crear una cuenta adicional para el bot, y dicha cuenta será señalada como bot (Fandom user rights).

Nótese que en WikiChron, tanto en la versión anterior a Monowiki, como en Monowiki, se efectúa un filtrado de bots durante el proceso de limpieza de los datos.

En este trabajo de fin de grado, resulta interesante estudiar qué tipos de usuarios existen en la comunidad de Fandom, puesto que, aunque en el conjunto de datos empleado en WikiChron, (tanto en su versión clásica como en la versión Monowiki), no se tiene acceso al grupo al que pertenecen los usuarios, en futuros análisis en los que se emplee la herramienta, se puede tratar de discernir a qué grupo pertenece un editor, atendiendo a criterios como el tipo de páginas que éste edita más.

Por otra parte, otro motivo por el que omitimos las clases de usuario definidas por la comunidad en WikiChron Monowiki es que la herramienta, desde la versión de WikiChron Classic, emplea dumps de MediaWiki, lo cuál no implica solamente a la comunidad de Fandom. Por ello, no es práctico llevar a cabo una diferenciación de los usuarios en función de su estatus, dado que éste es subjetivo para cada comunidad. Así, en WikiChron Monowiki, distinguimos entre usuarios anónimos y usuarios registrados, además de atendiendo a factores más complejos como su actividad o su antigüedad en la wiki.

2.4.2. Participación

Como se ha mencionado anteriormente, el análisis en comunidades colaborativas, se centra en proyectos exitosos como Wikipedia o Open Street Map, generando hipótesis en base a estos análisis y se corre el riesgo de generalizar los resultados obtenidos para todas las comunidades colaborativas.

No obstante, como parte del estudio de las comunidades colaborativas llevado a cabo en la FDI de la Universidad Complutense de Madrid, se han llevado a cabo análisis recientes, que se relacionan con el estudio de la participación en comunidades como Fandom, y algunos de los resultados obtenidos muestran que la realidad es otra.

El estudio de Serrano et al. (2018a), se basa en analizar la desigualdad en la participación que se da en comunidades colaborativas de tamaño y éxito medio, pertenecientes a Wikia, empleando la versión clásica de WikiChron.

En concreto, para llevar a cabo este estudio, se emplea una métrica que mide el nivel de desigualdad en una comunidad: el coeficiente de Gini. Este coeficiente mide la desigualdad o dispersión estadística, que, en el caso de WikiChron, está orientado a estudiar la desigualdad en el número de ediciones llevado a cabo por los usuarios. Para observar su evolución temporal, se hace uso del número de ediciones acumulado en cada mes (es decir, el total de ediciones en la wiki hasta cada mes).

Además del coeficiente de Gini, la medida de desigualdad por excelencia, se emplean otros ratios para arrojar luz sobre los resultados obtenidos con la métrica del coeficiente de Gini.

Se hace uso del ratio 10:90, que calcula el cociente entre el número de ediciones llevadas a cabo por el 10 % de contribuidores más activos, y el número de ediciones llevadas a cabo por el resto de usuarios en la wiki. Asimismo, se emplean otros ratios para estudiar la distribución de las ediciones entre el grupo de contribuidores más activos:

En primer lugar, se emplean tres ratios que calculan el cociente entre el contribuidor más prolífico, y los contribuidores en los percentiles 5, 10 y 20, en la lista de contribuidores ordenada de forma descendente según su número de contribuciones. En segundo lugar, se utilizan otros dos ratios: el primero calcula el cociente entre el número de contribuciones de los usuarios situados en los percentiles 5 y 10 respectivamente. El segundo ratio, calcula este mismo cociente, entre los usuarios situados en los percentiles 10 y 20.

Así, haciendo uso de estas métricas, se observa cómo se distribuye la participación en los primeros 118 meses de dos wikis de Wikia cuya temática son los hobbies: la wiki de Cocktails y la wiki de Home. La primera wiki tiene un total de 1700 artículos, y la segunda, un total de 900.

Observando el valor del coeficiente de Gini, se observa que ambas wikis siguen tendencias opuestas: la desigualdad de participación en la wiki de Cocktails crece, mientras que en la wiki de Home, decrece. Además, el

crecimiento de la desigualdad de participación en Cocktails se produce junto con el crecimiento de la wiki, mientras que en Home ocurre al contrario: la desigualdad de participación decrece a medida que la wiki va creciendo. Esta última observación contradice estudios que generalizan el hecho de que el crecimiento de una wiki y el aumento de la desigualdad de participación en la misma, están directamente relacionados.

Posteriormente, en este mismo estudio, se realiza un análisis de la participación de los usuarios en una wiki de preguntas y respuestas perteneciente a Wikia: la wiki de Wikianswers Pet. Los resultados de este análisis contradicen las afirmaciones realizadas en el estudio de las wikis de preguntas y respuestas exitosas como Yahoo! Respuestas, realizado en Khansa et al. (2015), en las que se sostiene que la participación en este tipo de wikis depende de factores como incentivos, el comportamiento pasado en la wiki, y estancia en la wiki (Serrano et al., 2018a).

Así, en sus resultados, se observa que, en la wiki de Wikianswers Pet, el número medio de ediciones por usuario es constante y tan sólo supera el 2,5 en pocas ocasiones. Por otra parte, se observa que el número de ediciones realizadas por usuarios anónimos crece hasta alcanzar el 51,5 % del total de ediciones en el periodo estudiado, y que el valor del ratio 10:90 es 1,3 y el coeficiente de Gini, decrece hasta tomar el valor de 0,53, siendo ambos valores sorprendentemente bajos. En definitiva, este artículo demuestra que la falta de estudios en comunidades menos exitosas, conlleva a una generalización de resultados obtenidos durante el análisis de proyectos más exitosos, que puede no ser acertada.

Capítulo 3

Herramientas

Para comprender en profundidad el contexto en el que se sitúa WikiChron Monowiki en cuanto a sus funcionalidades con respecto a otros productos existentes, se expone, a continuación, una síntesis de algunas de las herramientas de análisis de datos más prolíficas y utilizadas. En primer lugar, se exponen las funcionalidades principales de dos herramientas muy populares en el mundo empresarial de la visualización de datos: Tableau Software y Google Analytics. Posteriormente, nos centramos en tres herramientas que permiten visualizar diversas estadísticas sobre comunidades colaborativas en general: ohloh.net, y The Commoners Framework. Por último, exponemos información de tres herramientas de visualización, especialmente orientadas al estudio de wikis: WikiStats, WikiApiary y WAM de Wikia.

3.1. Herramientas genéricas para visualización de datos

En la ciencia de datos, y en especial, en el mundo empresarial, existen dos herramientas pioneras en visualización de datos. La primera es Tableau Software, una herramienta que en líneas generales, permite introducir un conjunto de datos en formato excel o .csv, que pueden ser extraídos de diversas fuentes, y pone en disposición del usuario una serie de funcionalidades, que le permiten visualizar los datos de varias maneras diferentes.

En segundo lugar, está Google Analytics, una herramienta orientada al análisis de mercado y de sitios Web, que crea los llamados reportes, en los que se analizan diferentes aspectos del sitio Web. A continuación, se proporciona una descripción detallada de ambos.

3.1.1. Tableau Software

Tableau es una plataforma de gran valor en el mundo empresarial, ya que permite llevar a cabo una visualización de datos enteramente interactiva. Su funcionamiento básico consiste en acceder a una fuente de datos (que se le puede proporcionar tanto a partir de la plataforma de Google Big Query, como a partir de MySQL, y numerosas fuentes de datos adicionales) (Murray, 2016). una vez disponible la fuente de datos, que puede estar disponible de dos maneras: live o extract (Medrano, 2016):

- Conexión live, o en vivo: al seleccionar el modo de conexión en vivo (o live) a la fuente de datos, siempre que se visualicen los gráficos que exponen los datos de la fuente, estos datos incluirán las últimas actualizaciones, ya que Tableau actualiza automáticamente la fuente de datos.
- Conexión extract, o en extracto: al seleccionar el modo de conexión en extracto (o extract), se deberá actualizar la fuente de datos de tableau manualmente para poder visualizar la última información obtenida.

En Tableau, se emplean diferentes tipos de gráficos para visualizar de la manera más sencilla posible, la información que se desee, a partir de la fuente de datos proporcionada. En la figura 3.1, se muestra una hoja de trabajo en esta herramienta.

En definitiva, es una herramienta que proporciona una gran experiencia de usuario, puesto que, una vez proporcionada la fuente de datos, el usuario tiene la opción de crear distintos tipos de gráficos: mapas de calor, gráficos de barras, gráficos de líneas, gráficos de tartas, y un largo etcétera. Tan sólo, debe colocar los campos del conjunto de datos (que pueden ser directamente provenientes del conjunto de datos original, o campos calculados a partir de varias columnas del conjunto de datos) que se desea visualizar en una de las múltiples dimensiones disponibles: filas, columnas, color, tamaño, texto, y un largo etcétera.

3.1.2. Google Analytics

Google Analytics es una herramienta Web, que permite observar diversos aspectos de un sitio Web. En concreto, es empleado por empresas de marketing, para la observación del progreso de campañas, en términos de número de usuarios que visitan el sitio Web, o número de usuarios que hacen click en un determinado espacio de la página Web. Así, algunas de las funcionalidades principales de Google Analytics son las siguientes:

- Cabe destacar que Google Analytics funciona al añadir un código JavaScript, que permite enviar datos de eventos a Google Analytics, Google Ads y Google Marketing Platform. Para instalar este fragmento de

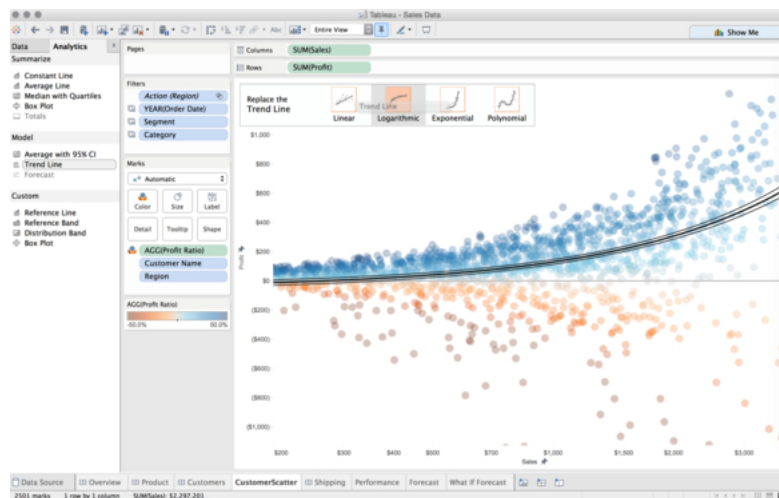


Figura 3.1: hoja de trabajo en Tableau

seguimiento global, se debe copiar el código que ofrece google analytics, haciendo los ajustes persistentes en él para indicar el sitio Web cuya información se quiere extraer, Clifton (2012).

- Google Analytics organiza la información obtenida del sitio Web en cuatro tipos de informes temáticos, Saleh (2016):
 - Informes de audiencia, que proporcionan información acerca de qué tipo de usuarios visitan el sitio Web.
 - Informes de adquisición, que proporcionan información acerca de la manera en la que los usuarios acceden al sitio Web.
 - Informes de comportamiento, que permiten observar estadísticas generales del sitio Web en cuanto a popularidad de las diferentes páginas Web.
 - Informes de conversión, que responden a preguntas como qué productos del sitio Web generan más compras, o qué trayectoria siguen los usuarios en el sitio Web antes de realizar una compra.
- Google Analytics permite filtrar datos que pueden contaminar la información, y perjudicar el análisis final. Es decir, funciona como una herramienta de limpieza de datos. Para filtrar estos datos, se indican las características de la información a incluir o a excluir, según la conveniencia (Drakhshandeh, 2015).
- Google Analytics permite definir eventos a analizar. La definición de eventos es un método disponible en el GATC, o Google Analytics Tracking Code, que es el código JavaScript mencionado anteriormente. Este método de definición de eventos, se emplea para registrar la interacción

de los usuarios con determinados elementos de la página web. Ejemplos de eventos son: "hacer click en un vídeo", "descargar un archivo", y un largo etcétera (Keyworth, 2017).

- Una funcionalidad de gran utilidad para el análisis de datos de Google Analytics, es que permite analizar la trayectoria seguida por los clientes dentro de un sitio Web. Para ello, se debe, en primer lugar, definir una serie de objetivos que el usuario ha de conseguir: hacer click en un determinado anuncio de la página Web, para posteriormente, hacer la compra de un producto (Han, 2019).

Así, este conjunto de funcionalidades, hace de Google Analytics, una de las herramientas más empleadas en el mundo de la visualización de datos en el área del marketing empresarial.

3.1.3. Conclusiones

La investigación y el análisis de estas herramientas nos ha permitido tener una visión más amplia del tipo de gráficos y opciones visualies que existen. De Tableau, hemos obtenido ideas como la posibilidad de incluir más de un tipo de visualización en WikiChron Monowiki. En concreto, la idea de incluir heatmaps la extraímos de esta aplicación. Por otra parte, la organización de la información que hace Google Analytics, en reportes de actividad, es algo similar a lo que se hace en WikiChron Monowiki, dividiendo los tipos de gráficos en categorías en función del tipo de información que se muestra en ellos (esto se encuentra así organizado desde su versión anterior en WikiChron Classic), por lo que con Google Analytics, vemos que la estrategia de organizar la información en categorías es efectiva y exitosa.

3.2. Herramientas centradas en Software Libre

Tras una exposición de dos de las herramientas de visualización de datos más empleadas en el mundo empresarial, se procede a proporcionar al lector un trasfondo teórico en cuanto a qué herramientas existen para visualizar la información existente en proyectos colaborativos de Software Libre. Se exponen dos herramientas orientadas a medir el valor de los productos generados en proyectos colaborativos. La primera herramienta de la que se habla, ohloh.net, está orientada al análisis de proyectos de Software libre, mientras que la segunda, The Commoners Framework, está orientada al estudio de todo tipo de proyectos colaborativos.

3.2.1. Openhub

Antiguamente conocida como Ohloh.net, se trata de una aplicación web que permite medir el valor de diferentes productos de software libre, propor-

cionando una funcionalidad para contrastar y comparar diversos aspectos de varios de ellos simultáneamente.

Fue creada en 2004 por Jason Allen y Scott Collison, ex-directores de Microsoft. Su funcionamiento básico consiste en extraer los datos de los sistemas de control de versiones en los que se encuentra el código en cuestión, y calcular diferentes estadísticas sobre los datos almacenados.

La herramienta actúa sobre proyectos que se encuentran en multitud de repositorios FLOSS, como por ejemplo Github y Sourceforge, y permite medir aspectos como el número de commits, el número de líneas de código o el número de contribuidores. La utilidad de esta herramienta web es especialmente alta para los desarrolladores de software libre, ya que les permite observar cómo progresa su producto e incluso emplear las estadísticas obtenidas en campañas de marketing, Allaho y Lee (2013). No obstante, esta herramienta no emplea ningún gráfico que permita visualizar las estadísticas, tan sólo se trata de números.

General	Arduino	GnuPG	Wikia
Project Activity	Moderate Activity	High Activity	Very High Activity
Open Hub Data Quality	Updated 22 minutes ago	Updated about 9 hours ago	Updated about 5 hours ago
Homepage	www.arduino.cc	www.gnupg.org	www.wikia.com
Project License	Creative Commons AttributionShareAlike 2.5-deleted, gpl, lgpl	gpl	gpl
Estimated Cost	\$608,620	\$4,200,084	\$38,632,580

Figura 3.2: Información general de los proyectos a comparar en ohloh.net

All Time Statistics	Arduino	GnuPG	Wikia
Contributors (All Time) View as graph	63 developers	276 developers	607 developers
Commits (All Time) View as graph	8253 commits	6994 commits	155784 commits
Initial Commit	over 21 years ago	over 13 years ago	over 10 years ago
Most Recent Commit	2 days ago	5 days ago	1 day ago

Figura 3.3: Estadísticas generales de los proyectos a comparar en ohloh.net

3.3. Bitergia Cauldron

Bitergia es una empresa que se encuentra en el parque científico de la Universidad Carlos III de Madrid, y cuyo objetivo es analizar los datos de proyectos de software libre. Entre sus servicios, proporciona tableros de información y reportes para este tipo de proyectos. En concreto, su herramienta

Code Analysis

Mostly Written In	C	Java	PHP
Comments	Average	Average	Low
Lines of Code View as graph	290,657 lines	45,978 lines	2,422,598 lines

Figura 3.4: Estadísticas generales de código en los proyectos a comparar en ohloh.net



Figura 3.5: Reporte de actividad de la herramienta Cauldron desarrollada por Bitergia.

Cauldron es un servicio que permite realizar un análisis de la actividad de un usuario de GitHub. Entre los datos que permite observar, se encuentra el número de commits, issues y pull requests. En Cauldron, se agrupa la información en un reporte (figura 3.5) que contiene las siguientes secciones:

- Sección de Overview: expone una visión general de los repositorios del usuario.
- Sección de Git: en ella se muestra información acerca de los commits, como el autor o el momento en el que se llevó a cabo el mismo.
- Sección de Issues: muestra información acerca del momento de apertura y resolución de los mismos.
- Sección de Pull requests: muestra información relativa a los pull requests del repositorio.

3.3.1. Conclusiones

Con esta investigación, se observa que hay algunas herramientas disponibles para llevar a cabo un análisis de proyectos colaborativos en general. No obstante, el estudio que estas herramientas permiten es muy superficial, y se basa en estadísticas que no se muestran visualmente. Por tanto, el emplear estas herramientas para llevar a cabo un estudio en wikis, no aportaría

tanta información como la que se puede obtener empleando WikiChron en cualquiera de sus versiones.

3.4. Herramientas específicas para el estudio de wikis

A continuación, se presentan cinco herramientas cuyo objetivo principal es el estudio en wikis. Algunas de estas herramientas presentan funcionalidades similares a WikiChron Monowiki. No obstante, ninguna de ellas está dotada de la completitud que presenta WikiChron Monowiki, y casi todas ellas (a excepción de history flow), parecen estar orientadas a proporcionar una visión de la wiki a los usuarios o creadores de la misma, más que a favorecer un análisis científico e intensivo de las mismas.

3.4.1. WikiStats

WikiStats¹ es una herramienta proporcionada por Wikimedia Foundation, con el objetivo de proporcionar una visión de diversas estadísticas de las wikis creadas por Wikimedia.

La herramienta presenta un buscador, en el que se puede introducir el nombre de cualquier wiki perteneciente a Wikimedia, para posteriormente, mostrar las siguientes estadísticas, divididas en tres categorías. Nótese que las gráficas proporcionadas, pertenecen a la Wikipedia Francesa.

La categoría “reading”, o lecturas, engloba métricas que permiten observar estadísticas como la evolución temporal del número de aparatos que se encuentran conectados en una wiki en un momento dado (figura 3.6), la evolución temporal del número total de páginas en la wiki (figura 3.7), o el número de visitas totales por país, en forma de mapa del mundo en la que la intensidad de los colores indica mayor o menor número de visitas (figura 3.8):

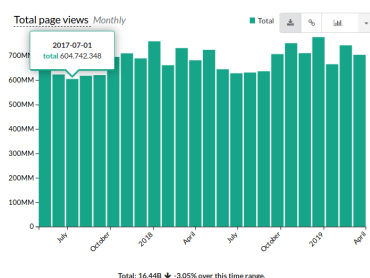


Figura 3.6: Métrica de aparatos únicos conectados en WikiStats.

¹<https://stats.wikimedia.org>

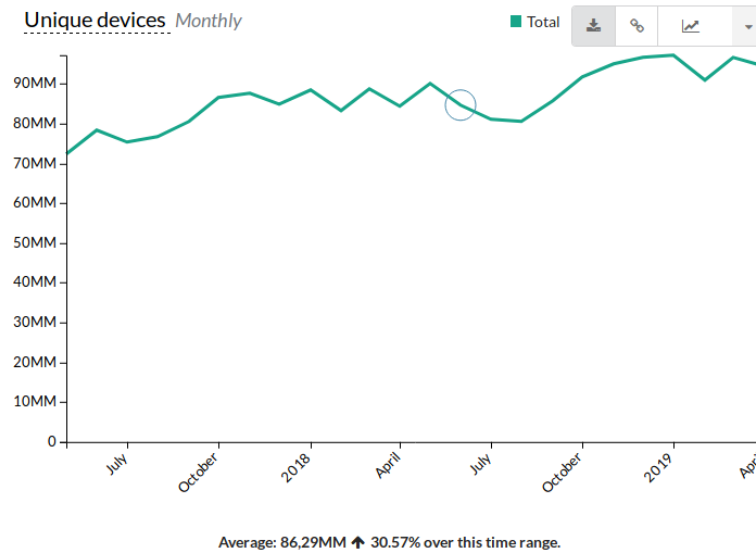


Figura 3.7: Métrica de número total de páginas visitadas en WikiStats.

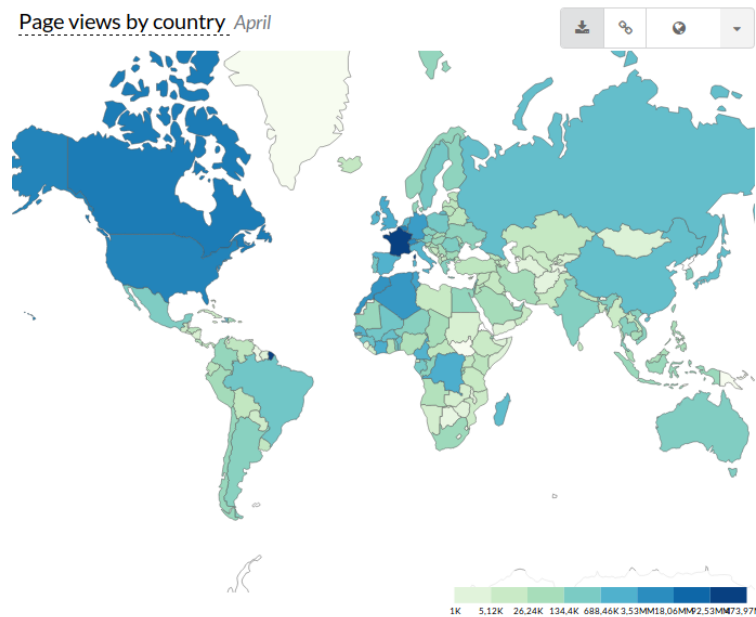


Figura 3.8: Métrica de número total de páginas visitadas por país en WikiStats.

La siguiente categoría es la categoría “contributing”, o contribución, que contiene métricas que permiten visualizar la evolución temporal de aspectos como el número de ediciones hechos por los usuarios, diferenciando

entre anónimos, registrados y bots (figura 3.9), el número de usuarios nuevos registrados (3.10) y el número de editores activos, con más de cinco ediciones (figura 3.11). Por último, WikiStats presenta la categoría “content”, o

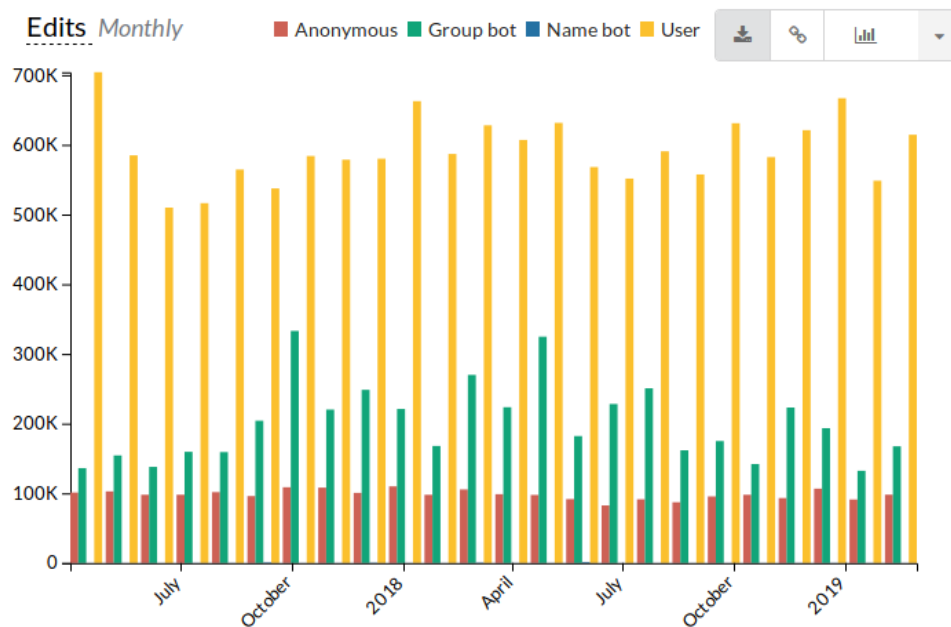


Figura 3.9: Métrica de ediciones en WikiStats.

contenido, que engloba métricas que permiten estudiar la evolución temporal de aspectos como el número de ediciones en páginas (figura 3.12), una métrica que muestra la suma de las diferencias en bytes llevada a cabo en cada edición (figura 3.13), y por último, una métrica que muestra la suma de la diferencia absoluta en bytes llevada a cabo en cada edición (figura 3.14).

3.4.2. WAM de Wikia

El WAM de Wikia² es una herramienta Web, proporcionada por la propia comunidad de Wikia (actualmente llamada Fandom), que definen como “un indicador de la fuerza y del impulso de una comunidad de Fandom”.

Como se observa en la figura 3.15, el WAM de Wikia proporciona dos rankings:

Un ranking general, es decir, calculado sobre todas las wikis pertenecientes a cualquier categoría, y **un ranking vertical**, calculado para todas las wikis pertenecientes a una misma categoría. A su vez, proporciona una **puntuación de 0 a 100**.

No obstante, los aspectos específicos del cálculo del ranking no son pro-

²<https://community.fandom.com/wiki/WAM>

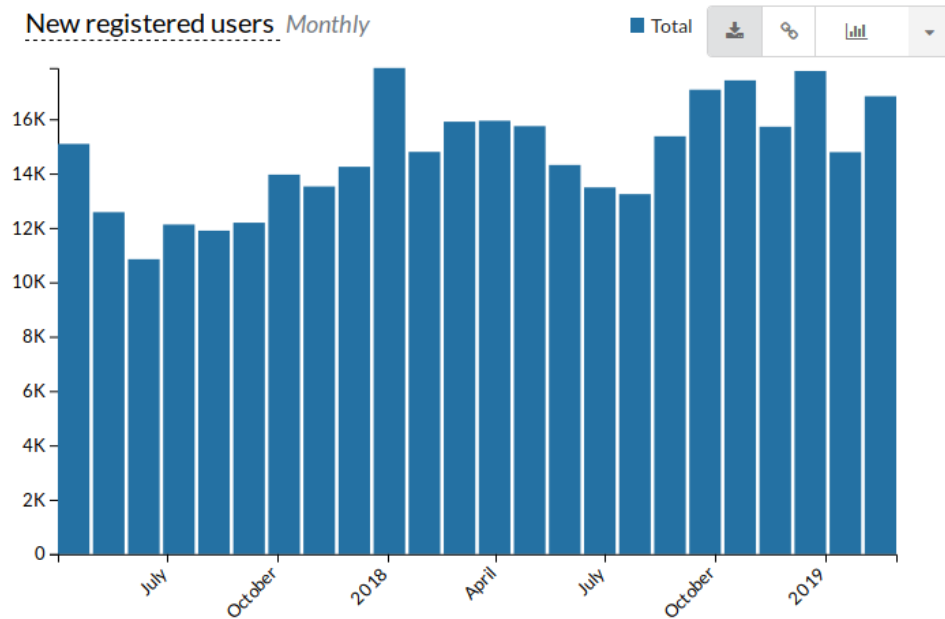


Figura 3.10: Métrica de usuarios nuevos registrados en WikiStats.

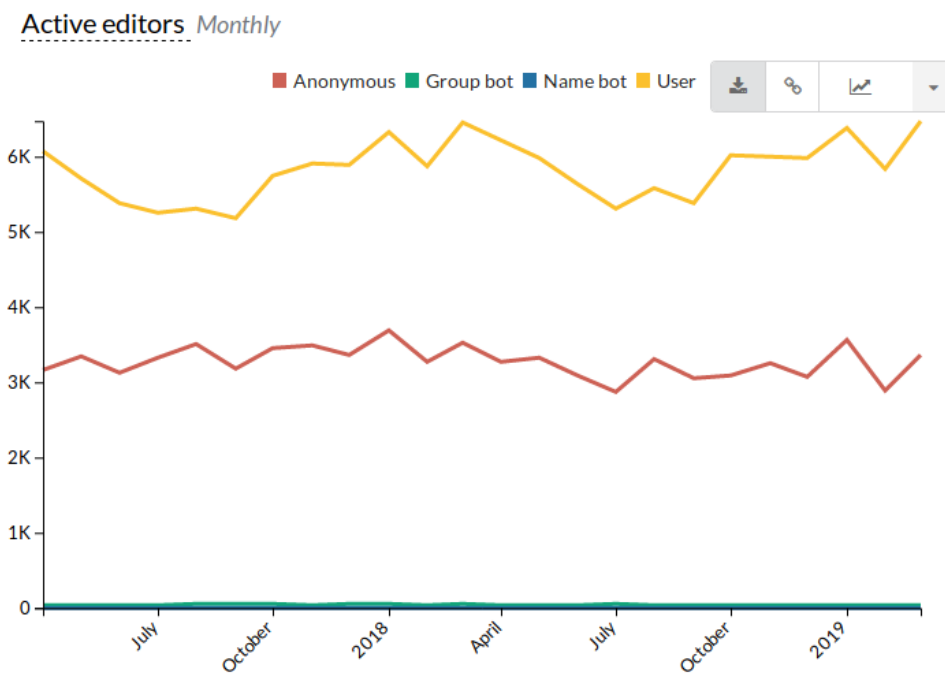


Figura 3.11: Métrica de usuarios nuevos registrados en WikiStats.

porcionados como información pública por Wikia, dado que quieren evitar que las comunidades, siendo conscientes de ello, traten de manipular el ran-

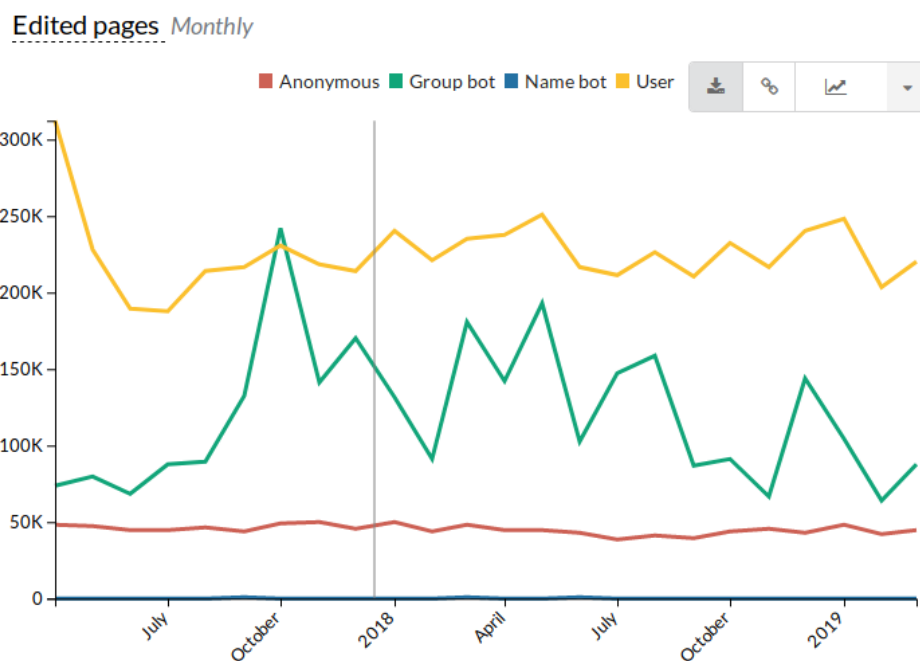


Figura 3.12: Métrica de páginas editadas en WikiStats.

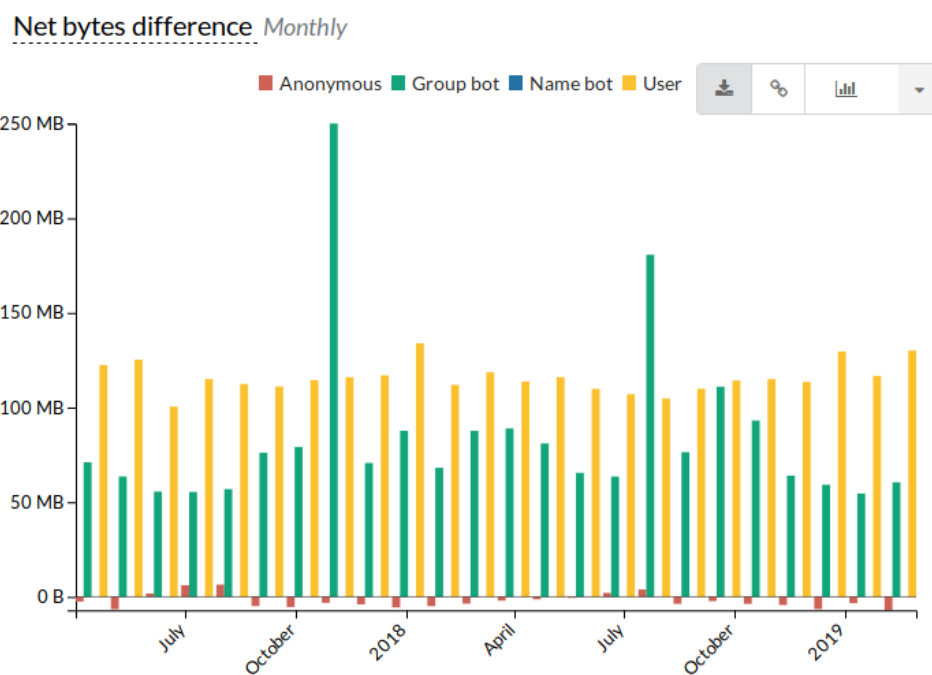


Figura 3.13: Métrica de diferencia neta en bytes en WikiStats.

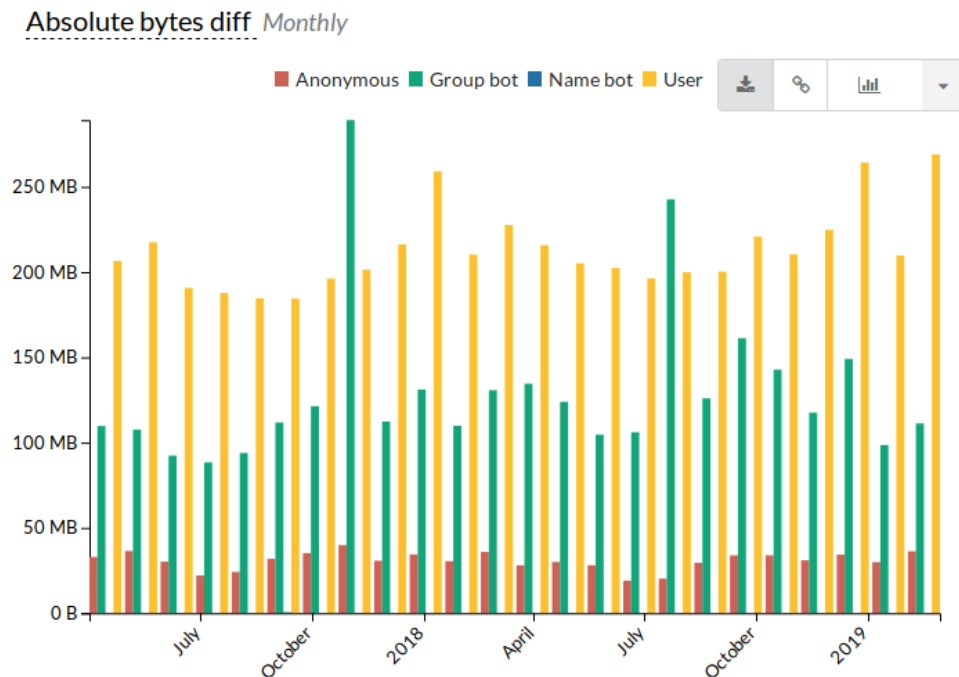


Figura 3.14: Métrica de diferencia absoluta en bytes en WikiStats.

A-Z	TV	GAMES	BOOKS	COMICS	LIFESTYLE	MUSIC	MOVIES
ALL							

Language	All	Date	February 1, 2016	FLIPLINE	Q
----------	-----	------	------------------	----------	---

Rank	WAM Score	Peak Rank	Wikia URL	Vertical	Vertical Rank	Admins
278	↑ 92.90	74	fliplinestudios.wikia.com	Games	141	
2564	↑ 44.23	85	es.flipline-studios-espanol.wikia.com	Games	1104	
4185	↑ 14.39	171	fliplinefancition.wikia.com	Lifestyle	363	

Figura 3.15: WAM de Wikia.

king (Fandom frequently asked questions). La única información que proporcionan es que atienden a aspectos como el tráfico, el compromiso de los usuarios para con la wiki, y su crecimiento.

Así mismo, alegan que el ranking de WAM es una buena herramienta para administradores, fundadores y contribuidores para comprender cómo evoluciona su wiki.

En líneas generales, el WAM no se considera una herramienta válida para el análisis en profundidad de las wikis de Wikia, dado que no se proporcionan

los detalles de la implementación. No obstante, es útil como herramienta de motivación para los usuarios comprometidos con su wiki, y quieran verla crecer.

3.4.3. WikiApiary

WikiApiary³ nació el 21 de Diciembre de 2012, y es una wiki que monitoriza a todas las wikis que funcionan bajo el software de MediaWiki.

Una vez que un sitio Web está registrado en WikiApiary, una colección de bots comenzará a almacenar información sobre la actividad en dicha wiki. Entre los datos almacenados se encuentran:

- La versión de software de MediaWiki utilizada.
- La actividad de la wiki en términos de ediciones.
- El uso de Semantic MediaWiki.

WikiApiary muestra unas estadísticas generales sobre todas las wikis que emplean MediaWiki y están registradas en WikiApiary, entre las que se incluyen el número de wikis registradas en WikiApiary, el número total de páginas y de ediciones en la wiki, o el número total de usuarios.

Además de estas estadísticas generales, WikiApiary proporciona una colección de las wikis activas en la herramienta, categorizadas en función del número de páginas de las mismas.

3.4.4. Conclusiones

Con la investigación anterior, hemos descubierto algunas herramientas que permiten estudiar algunos aspectos en wikis. Al igual que WikiChron en sus versiones Classic y Monowiki, la herramienta de WikiStats proporciona varios gráficos de evolución temporal. En algunas de las gráficas que se muestran (como por ejemplo, en el número de aparatos conectados a una wiki, o el número de visitas hechas desde diferentes países), se emplean datos que no se encuentran disponibles en el dump de MediaWiki, por lo que no es posible incluir dichas visualizaciones en WikiChron.

No obstante, la mayor parte de las otras métricas que se presentan, también se pueden observar con WikiChron, en sus versiones Classic o Monowiki. por otra parte, WikiStats lleva a cabo una categorización de los usuarios en clases, al igual que lo que vamos a hacer en WikiChron Monowiki. No obstante, la clasificación que presentamos en nuestra herramienta, proporcionará detalles más profundos acerca de la estructura de la comunidad, dividiendo a los usuarios en clases atendiendo a aspectos más complejos que el estar registrado o no, como por ejemplo, la antigüedad editando en la wiki.

³https://wikiapiary.com/wiki/Main_Page

por otra parte, la utilidad del Wam de Wikia y de WikiApiary está orientada a dar feedback a la propia comunidad, y no a su uso para estudiar sus características en profundidad.

3.5. WikiChron

WikiChron es una nueva herramienta web para el análisis y visualización de la evolución de las wikis. Procesa los volcados históricos de wikis de mediawiki, calcula diferentes métricas en estos datos y los grafica en gráficos interactivos. Un ejemplo de métrica que ya venía implementada en WikiChron es “Active Users”, que calcula sobre los datos de la wiki cuantos usuarios tienen un mínimo de una edición en un mes.

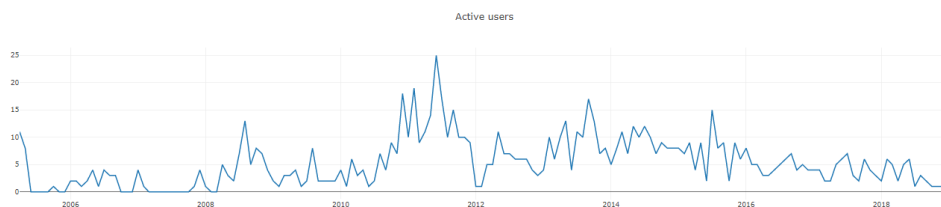


Figura 3.16: Active Users

Permite comparar diferentes wikis en los mismos gráficos, esta herramienta ha sido creada por el grupo de trabajo GRASIA (Grupo de investigación en Aplicaciones Sociales e Interdisciplinarias basadas en Agentes) cuyo objetivo es ayudar a las comunidades a realizar su potencia a través de la ingeniería de sistemas socio-técnicos (<https://grasia.fdi.ucm.es/newmain/language/es/>).

3.5.1. Tecnologías.

En WikiChron se emplean las siguientes tecnologías:

- **Python:** de entre todas las librerías que posee este lenguaje, caben destacar las dos más importantes dentro de WikiChron: Pandas y Plotly.
 - Pandas: es una biblioteca software escrita como extensión de NumPy y ofrece estructuras para manipular tablas numéricas y series temporales, es apropiada para el análisis de datos.
 - Plotly: proporciona herramientas de visualización.
- **Dash:** es un framework de Python para crear aplicaciones web. Permite componer elementos de interfaz de usuario, con menús desplegables, controles deslizantes y gráficos. La ventaja es que no requiere javascript.

- **Flask:** permite crear aplicaciones web rápidamente y con un mínimo número de líneas de código.
- **XML:** es un lenguaje de marcado parecido a HTML y también sirve como intercambio de información, al igual que JSON. En WikiChron, se usa de forma indirecta, solo está en este formato los datos descargados de las wikis (este proceso se explicará más adelante en la sección de recogida y formato de datos).
- **GitHub:** es una plataforma de desarrollo colectivo para alojar proyectos usando el sistema de control de versiones de Git. WikiChron está en esta plataforma y la URL es: <https://github.com/Grasia/WikiChron>

3.5.2. Formato de datos

Los datos se guardan en archivos csv. Cada uno de estos archivos recoge la información de una wiki entera y cada línea del csv indica una contribución que contiene la siguiente información:

- **Page_id:** el identificador de la página a la que se ha contribuido dentro de esa wiki.
- **Page_title:** el título de la página que se ha editado.
- **Page_ns:** el tipo de la página que se ha editado, si es un artículo o una talk page, por ejemplo.
- **Revision_id**
- **Timestamp:** indica la fecha y la hora exacta de la edición.
- **Contributor_id:** el identificador que se le asigna a un usuario cuando se registra, si es un usuario anónimo, aparecería la IP del dispositivo con el que se ha conectado.
- **Contributor_name:** el nombre del usuario, si no está registrado, aparece como “Anonymous”
- **Bytes:** La cantidad de bytes que tiene esa página tras la edición.

3.5.3. Estructura de la APP

3.5.3.1. Versión 1:



Figura 3.17: Página principal WikiChron V1

La figura 3.17 es la pagina principal de WikiChron v1 la barra azul mostrada en la parte izquierda de la pantalla es el “side bar”, su función principal es la elección de distintas wikis y métricas para la generación de gráficos.

Tiene dos pestañas wikis y metrics, en la pestaña wikis podemos seleccionar varias wikis para su comparación y en la pestaña métricas tenemos varias pestañas que nos indican las categorías de las métricas que podemos seleccionar para su calculo sobre las wikis seleccionadas.

Dentro de las métricas podemos destacar varias categorías según en lo que se centren:

- **Pages:** se centran en las páginas y los artículos.
- **Edits:** se centran en las ediciones.
- **Users:** se centran en los usuarios.
- **Ratios:** relaciona por ejemplo ediciones o artículos con usuarios.
- **Distribution of Participation:** introduce el coeficiente de Gini, explicado previamente.

Una vez hayamos seleccionado como mínimo una wiki y una métrica se desbloquea el botón compare, si clicamos en este botón se nos bloqueará la pantalla mientras se están generando los gráficos.

Al generar los gráficos se nos mostraran en la pantalla principal, en la parte superior, las wikis y las métricas que hayamos seleccionado, el “time

axis ” nos permite marcar si el gráfico lo queremos en meses naturales o meses desde el nacimiento de la wiki. Asi como tenemos el “date slider” que nos permite elegir los meses entre los que queremos representar la wiki siempre teniendo como límite el principio y el fin de la wiki mayor seleccionada.

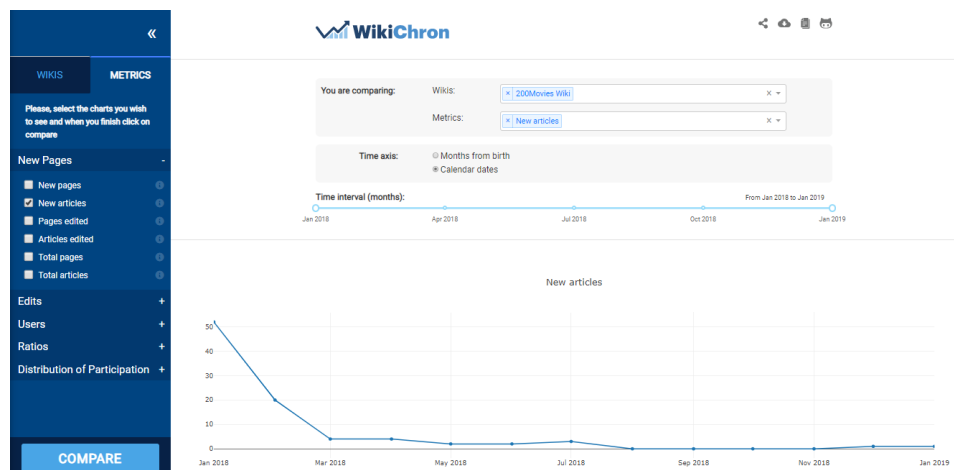


Figura 3.18: Gráfico con una wiki seleccionada

A continuación mostramos un ejemplo en que seleccionamos meses desde el nacimiento de la wiki y acortamos los meses en los que se representa la gráfica:

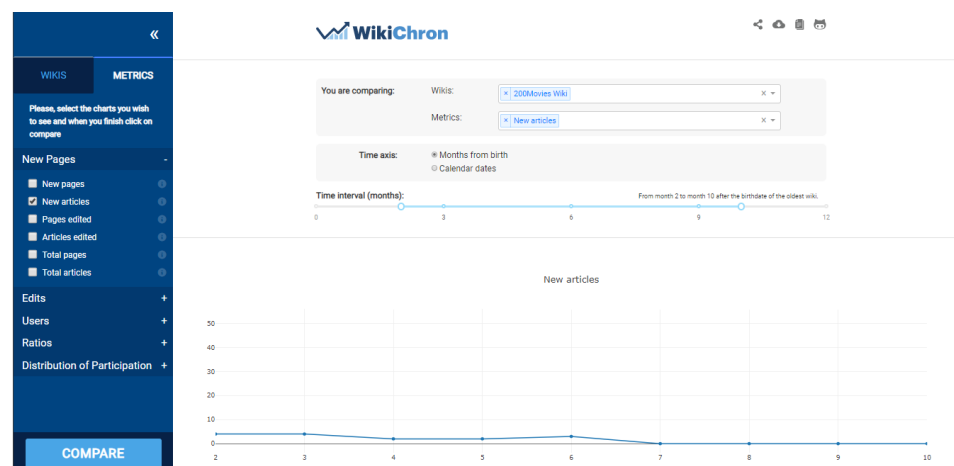


Figura 3.19: Gráfico acortando el date slider y seleccionando meses desde el nacimiento

En el caso de seleccionar varias métricas y varias wikis, se generan tantas gráficas como métricas hayamos seleccionado y en cada gráfico se calcula la métrica sobre todas las wikis seleccionadas como se muestra en la siguiente

imagen.

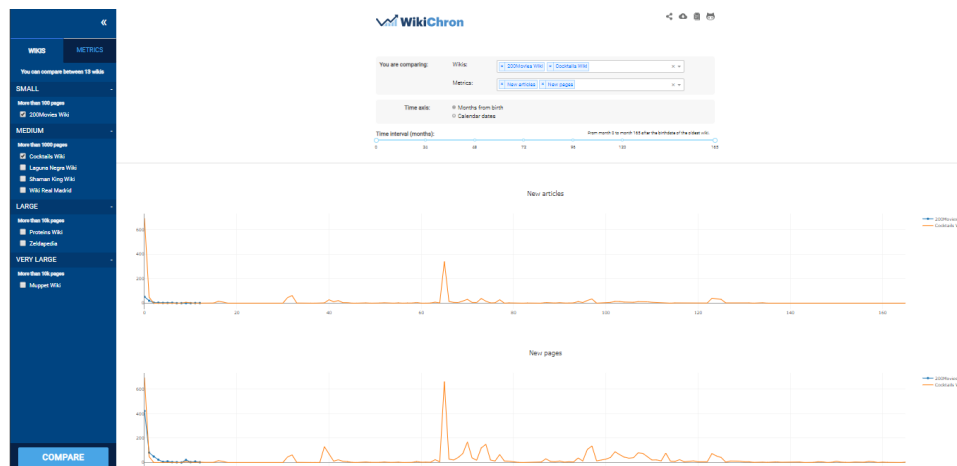


Figura 3.20: Selección de varias 2 métricas y 2 wikis

Una vez enseñado el front-end de wikichron v1, hemos realizado un diagrama de como se organiza internamente la aplicación y vamos a resumir los archivos que consideramos de mayor importancia:

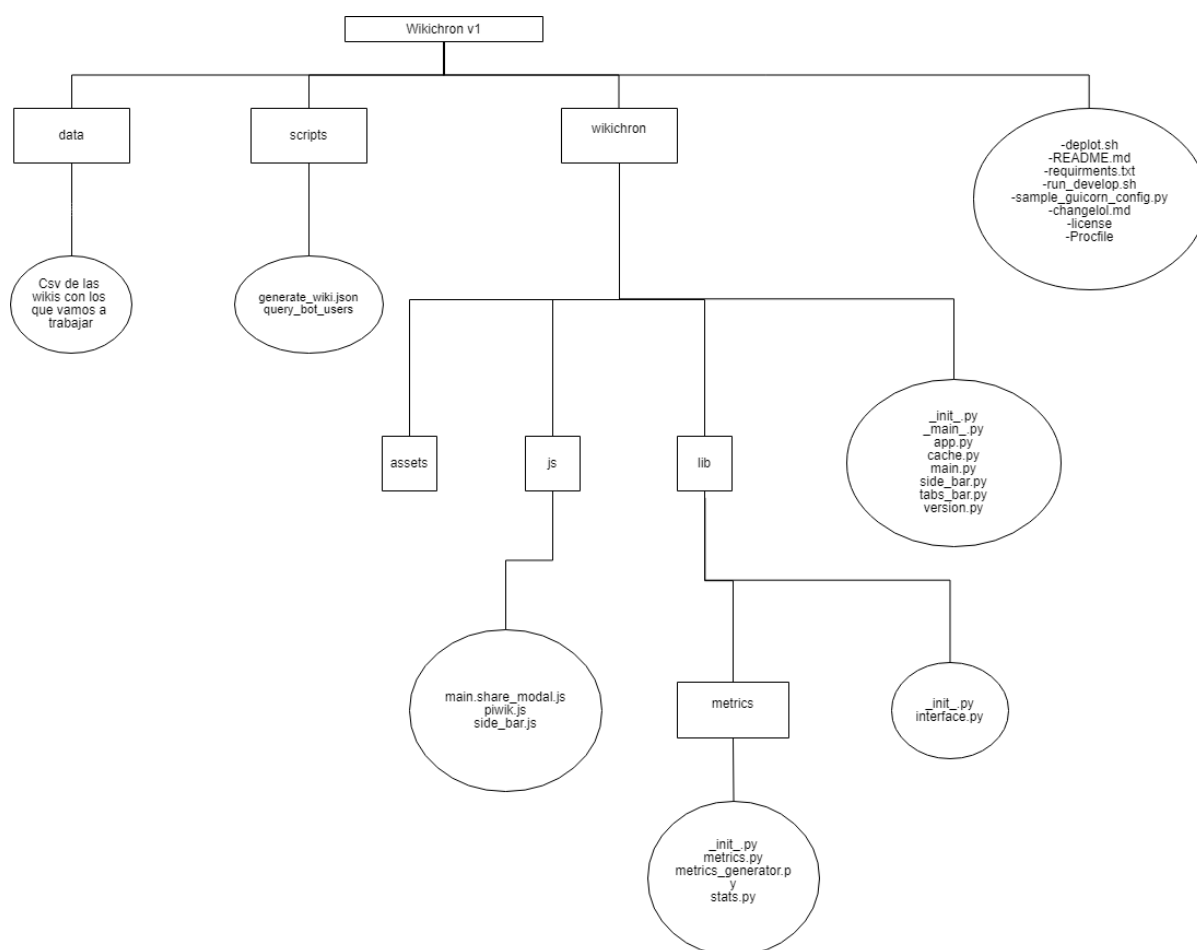


Figura 3.21: Diagrama Wikichron v1

Back-end

- **Wikis.json:** en este archivo se incluyen los metadatos de las wikis que vamos a cargar en wikichron, para poder trabajar con ellas.
- **Main.py:** este archivo controla la generación de gráficos, el time axis, el date slider y mantiene actualizado los gráficos ante cualquier cambio del usuario.
- **Interface.py:** se encarga de cargar los datos y prepararlos, es decir, transformar los datos que están en archivos csv en dataframe de pandas y elimina los bots, y ordenarlos por fecha.
- **Metrics_generator.py:** en este archivo se declara la cabecera de todas la métricas disponibles y se genera un diccionario que contiene como clave un código único de métrica y como valor la métrica que corresponde a ese código.

- **Stats.py:** en este archivo se define cada métrica, cada métrica devuelve una serie que es la que luego se utiliza para generar la gráfica.

Front-end

- **Run_develop.sh:** es una script que se incluye en los últimas actualizaciones de wikichron v1, para arrancar la aplicacion directamente.
- **Tabs_bar.py:** define la barra superior que contiene los botones de compartir, descargar los datos calculados, acceso directo a la documentación y acceso directo a github.
- **Side_bar.py:** define la barra de la izquierda, en la que se seleccionan las wikis y las métricas y se habilita el botón compare cuando mínimo han sido seleccionadas una wiki y una métrica.
- **App.py:** este archivo es el que genera la aplicación, se encarga de mostrarnos la página de inicio de wikichron, incluyendo el “side bar” aunque este es definido en otro archivo y comprueba las métricas y wikis que están disponibles.

3.5.3.2. Versión 2:

Esta nueva versión se publicó (en Github) y se desplegó en Abril de este año (2019). Plantea unas cuentas diferencias con respecto a la versión anterior, como la aparición de una tecnología nueva denominada Flask y algunos cambios más en la vista de la aplicación. Esta versión surge de la necesidad de integrar dos aplicaciones más, una de ellas es monowiki (nuestro TFG) y otra networks (otro TFG), todas ellas tienen en común la pantalla de selección de las wikis y además se crea otra pantalla adaptada para seleccionar las métricas o las redes según la aplicación en la que estemos. De esta forma es bastante más fácil introducir nuevas funcionalidades a la aplicación. La pantalla principal la podemos ver de la siguiente forma:



Figura 3.22: Página principal WikiChron V2

En la pestaña de WikiChron compare está el wikichron de la versión 1 y las otras dos corresponden a las nuevas aplicaciones que hemos dicho antes que se quería integrar en WikiChron. Si pasamos a la siguiente pantalla de WikiChron compare veremos lo siguiente:

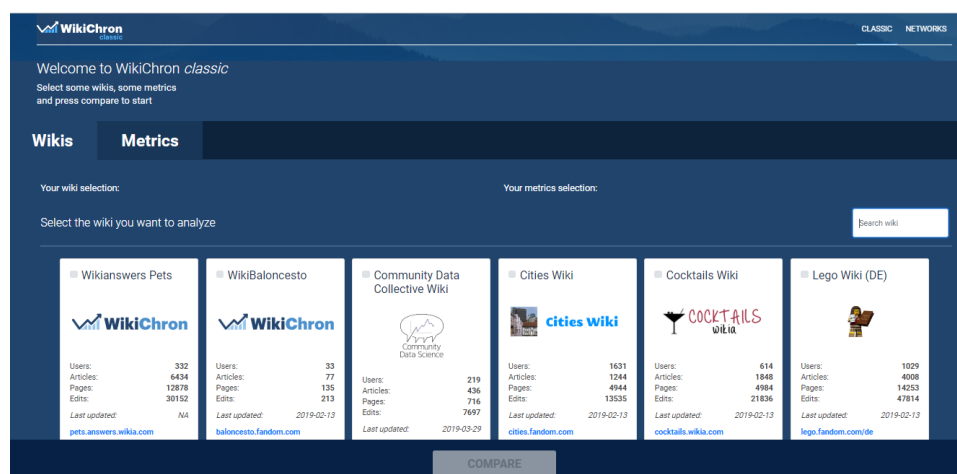


Figura 3.23: Pantalla de selección de Wikis.

Como podemos ver, se ha quitado el menú de selección que había en la parte izquierda en la anterior versión y se ha hecho una vista a parte para seleccionar las wikis y otra para seleccionar las métricas.

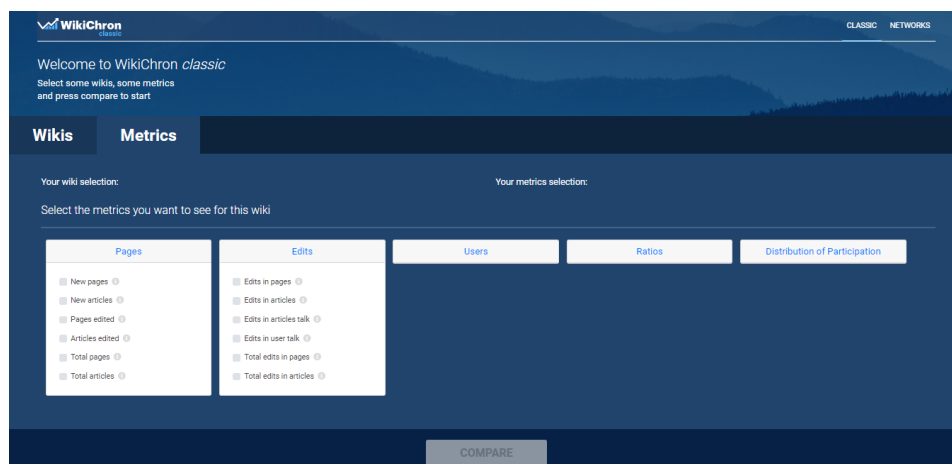


Figura 3.24: Pantalla de selección de métricas.

Aquí vemos como en la pantalla de selección de métricas aparecen divididas por categorías, al igual que en la versión anterior y al pinchar dentro de cada categoría aparecen las métricas asociadas a ella.

Una vez visto esto, si seleccionamos alguna wiki y alguna métrica y le damos al botón de “compare” que aparece en la parte inferior de la pantalla, aparecería la siguiente vista:



Figura 3.25: Pantalla donde aparecen los gráficos.

Esta pantalla es muy parecida a la de la versión 1 quitándole el menú de selección que había a la izquierda. En cuando al manejo de las opciones, es exactamente igual a lo explicado anteriormente en la versión 1.

Hemos hablado de los cambios visuales de esta versión y ahora vamos a explicar los cambios en la estructura interna de la aplicación.

Lo primero decir que WikiChron v1, en esta versión está dentro de: "WikiChron/wikichron/dash/apps/classic/", el contenido es casi el mismo aunque

se ha reestructurado la información. hemos hecho un diagrama que representaría, a grandes rasgos, la estructura interna de la aplicación (solo para WikiChron Compare).

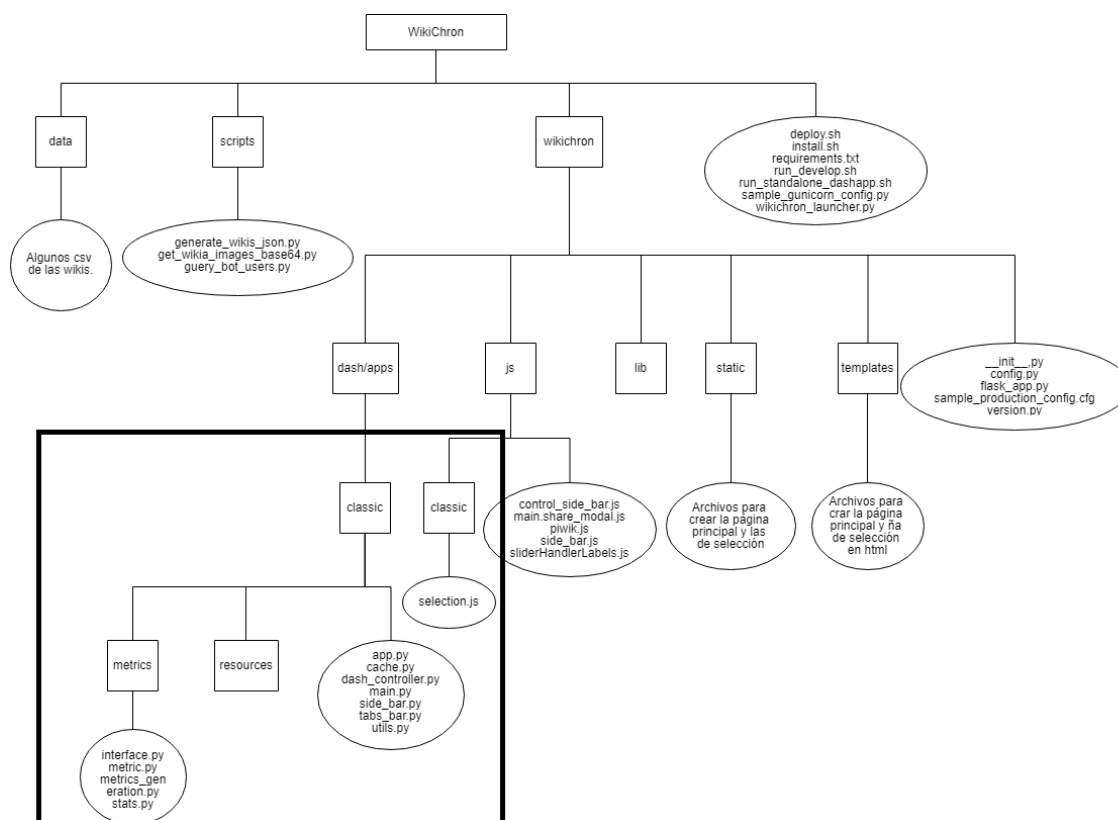


Figura 3.26: Diagrama estructura WikiChron V2

Se puede observar en los gráficos 3.21 y 3.26 como los cambios entre la versión uno y la versión dos se basan en separar la parte común de la aplicación, como por ejemplo, la selección de las wikis, de la parte específica de cada una de las subaplicaciones. El cuadro negro de la segunda imagen incluye el nuevo esquema que tiene cada modo dentro de WikiChron.

Capítulo 4

Contribuciones en WikiChron Classic

En esta sección vamos a hablar de algunas contribuciones que hemos realizado a la herramienta de WikiChron Classic. En todas estas contribuciones hemos seguido el patrón principal, es decir, las métricas se representan en series temporales, donde el eje de abscisas indica el tiempo en meses y el eje de ordenadas, el valor de la serie temporal para cada mes.

El objetivo de realizar estas contribuciones, por un lado, fue entender como se calculaban y había que desarrollar las métricas, y por otro, familiarizarse con la arquitectura de WikiChron para poder replicar aquellas partes necesarias para crear nuestra propia herramienta. Además, todas ellas estaban implementadas como issues abiertos en GitHub y se han realizado los correspondientes pull request en el repositorio oficial de la herramienta.

4.1. Usuarios activos registrados y anónimos

Estas métricas definen a usuarios, en el primer caso solo registrados y en el segundo caso solo anónimos, y ambos tiene que tener como mínimo una edición en ese mes. En este caso, el eje de ordenadas representa el número de usuarios que han editado ese mes. Para llevar a cabo la implementación de estas dos métricas, hemos creado una función auxiliar que devuelve el número de usuarios que han hecho más de un número determinado de ediciones por cada mes, y, es usada también en el desarrollo de alguna otra métrica explicada posteriormente.

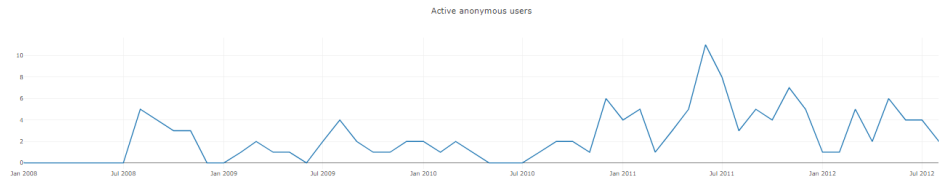


Figura 4.1: Métrica de usuarios activos anónimos, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.

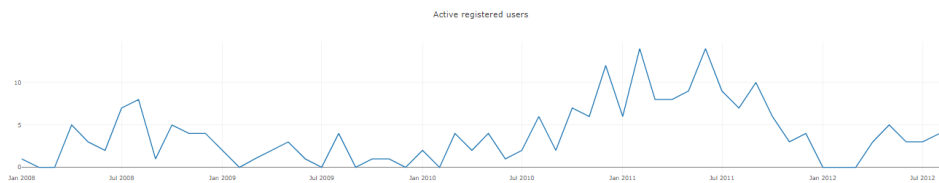


Figura 4.2: Métrica de usuarios activos registrados, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.

Unos ejemplos de visualización de estas métricas para la wiki de Cocktails se muestra en las figuras 4.1, 4.2

4.2. Usuarios activos con más de 4, 24 o 99 ediciones

Estas tres métricas son parecidas a la explicada en el punto 4.1, pero cambia en el número mínimo de ediciones que se deben de realizar al mes (4, 24 o 99 respectivamente). El propósito es determinar el trabajo realizado en una wiki por los usuarios en cada mes. Para realizar la implementación de estas métricas hemos usado la función auxiliar mencionada en el punto anterior (4.1).

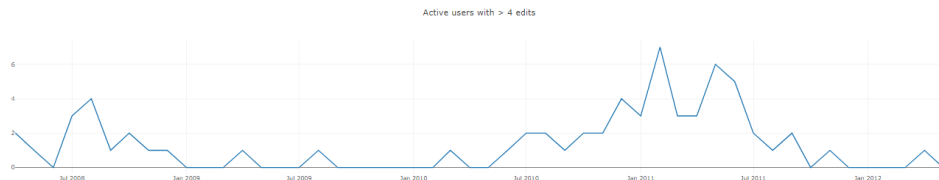


Figura 4.3: Métrica de usuarios activos con más de 4 ediciones, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.

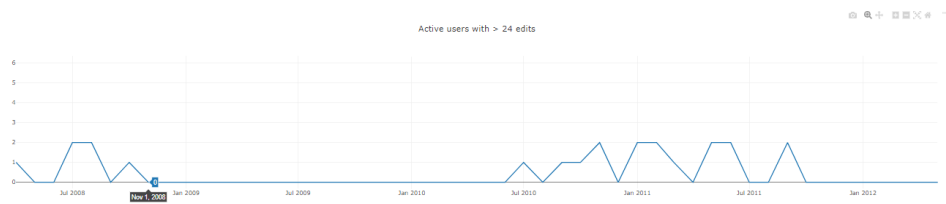


Figura 4.4: Métrica de usuarios activos con más de 24 ediciones, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.

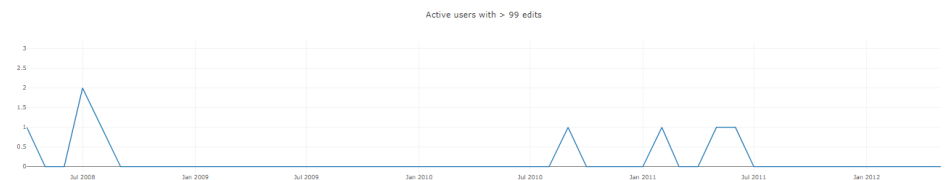


Figura 4.5: Métrica de usuarios activos con más de 99 ediciones, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.

Unos ejemplos de visualización de estas métricas para la wiki de Cocktails se muestra en las figuras 4.3, 4.4 y 4.5

4.3. Wiki durmiente

Esta métrica se encuentra en wikichron classic, muestra en que meses una wiki esta activa o esta dormida y esto nos permite tener claro os periodos en los que la wiki no está activa y poder analizar las causas. Una wiki dormida es aquella que lleva 3 meses seguidos sin ninguna edición tomamos esta definición del siguiente paper Zhu et al. (2014). En este caso, el eje de ordenadas tendria unicamente dos valores, 1 en el caso en el que la wiki este dormida y 0 en caso contrario.

Uno ejemplo de visualización de est métricas para la wiki de Cocktails se muestra en la figura 4.6.

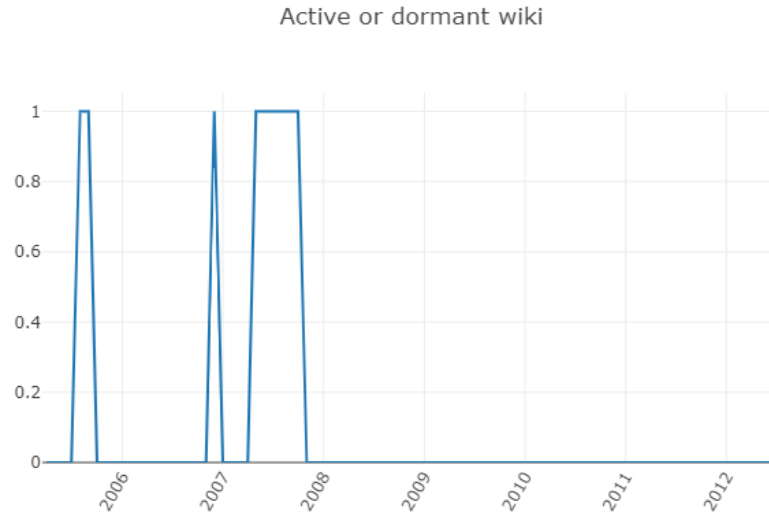


Figura 4.6: Métrica de wiki durmiente, en la wiki de Cocktails, durante el periodo temporal de abril de 2005 a julio de 2012.

4.4. Métricas de retención de usuarios de Wikimedia

Las métricas explicadas a continuación sirven para medir la reención de usuarios en la wiki, es decir, el número de usuarios que una vez realizada la primera edición continúan editando.

4.4.1. Nuevo editor reincidente

Esta métrica consiste en visualizar la cantidad de usuarios que, dentro de los primeros siete días desde su registro, realizan dos ediciones de sesión¹. Esta métrica sirve para tener una tasa de usuarios que vuelven, ya que no solo es importante atraer a nuevos usuarios, sino que también hay que retenerlos y está basada en el siguiente artículo de wikimedia: Returning

A la hora de la implementación hemos tenido en cuenta que podemos tener al mismo usuario en dos meses distintos, este es el caso en el que el usuario se haya registrado al final de un mes y los siete días siguientes se solapan con el mes posterior. Por eso, solo le contamos como usuario “reincidente” en el primer mes y no en el siguiente.

¹Tiene que haber una diferencia mínima de 60 minutos entre una edición y otra

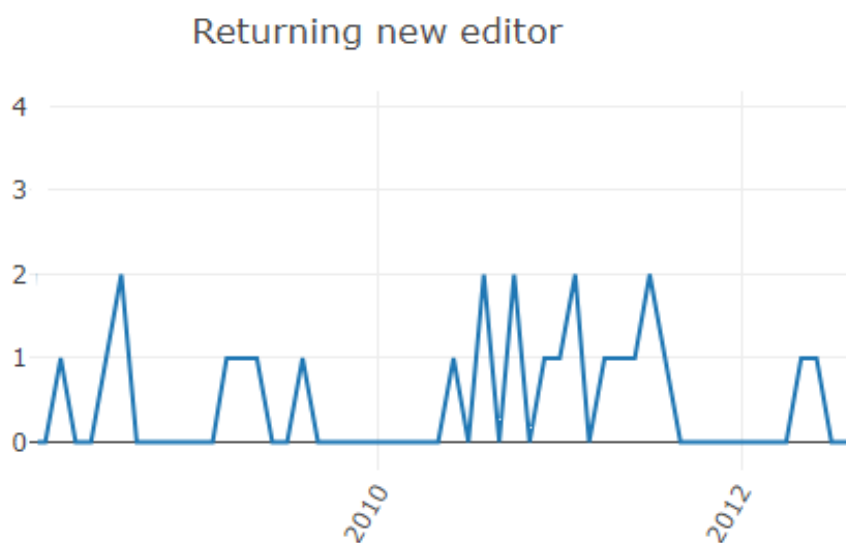


Figura 4.7: Métrica de Returning new editor, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.

Un ejemplo de visualización de esta función, para wiki de Cocktails, se muestra en la figura 4.7.

4.4.2. Nuevo editor superviviente

Esta métrica pretende visualizar la cantidad de usuarios que realiza al menos una edición entre las 24 horas después de su registro y que a demás, en el segundo mes desde que se registro, hace al menos otra edición más. Permite medir la retención de usuarios en un periodo de tiempo más amplio que la métrica anterior y nos hemos basado en el siguiente artículo de wikimedia: Surviving.

A la hora de la implementación hemos tenido en cuenta que el segundo mes pille una parte de un mes y otra parte de otro. En este caso, cogemos el mes más elevado. Por ejemplo, si un contribuidor se registró el día 20 de abril, el segundo mes tras esa fecha estaría comprendido entre mayo y junio y si ha hecho una edición en cada uno de esos meses, vamos a tener a este usuario repetido, por lo tanto lo que queremos es contarle solo en el máximo mes que es contado como superviviente (junio).

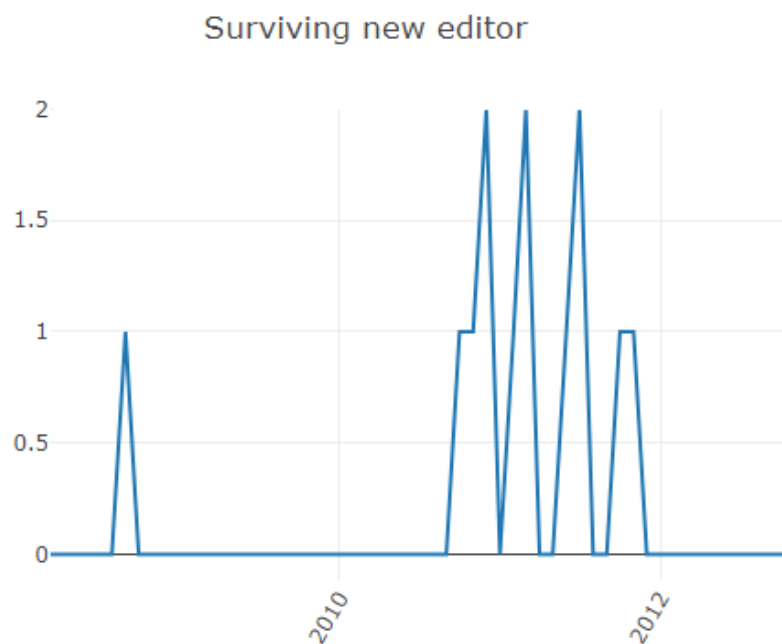


Figura 4.8: Métrica de Surviving new editor, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.

Un ejemplo de visualización de esta función, para wiki de Cocktails, se muestra en la figura 4.8.

WikiChron Monowiki

5.1. Motivaciones

WikiChron Monowiki surge de la necesidad de observar características que no es posible mostrar de forma clara para varias wikis a la vez. Entre estas características, destacan las siguientes:

- Analizar a los usuarios activos en cada mes en una sola comunidad, atendiendo a una serie de criterios, como la experiencia en número de ediciones, basándonos en el estudio realizado por Kittur et al. (2006) para Wikipedia, u otros criterios como la antigüedad del usuario en la wiki, la fecha de su última contribución en la misma, o el número de meses seguidos que lleva editando.
- Analizar como se distribuyen las ediciones entre las diferentes categorías de usuarios descritas anteriormente.
- Observar cómo se distribuyen los editores y las ediciones que se realizan en un mes en las páginas la wiki con el fin de ver si se concentran en unas pocas páginas o no.
- Determinar qué tipos de usuarios añaden o borran más conocimiento en la wiki, para discernir si se trata de usuarios nuevos, usuarios pertenecientes a la élite de la wiki en cuanto a su participación, usuarios con mucha antigüedad, etcétera, y observar la evolución temporal del conocimiento existente en la wiki.

Es importante tener en cuenta que los cálculos realizados para todas estas métricas, son sin contar a los usuarios anónimos, ya que estos usuarios se identifican por la dirección IP, y este identificador no es fiable, dado que desde un mismo dispositivo se pueden conectar varios usuarios y un usuario se puede conectar desde varios dispositivos. Además, según nuestro criterio, el

usuario registrado muestra un deseo de pertenecer a la comunidad, registrándose y usando un nombre que le da una identidad dentro de ella, mientras que un usuario anónimo difícilmente puede considerarse como miembro de la comunidad.

En definitiva, con WikiChron Monowiki se pretende permitir que los estudiosos en el área de investigación de las comunidades colaborativas, puedan observar elementos de una sola wiki en profundidad, sin centrarse en la comparación, sino en cómo evolucionan estos elementos y cómo se relacionan con otros elementos de la misma wiki.

A continuación, tras explicar los cambios hechos en WikiChron Classic para permitir distintos tipos de visualización, proporcionamos una descripción detallada de una amplia serie de métricas que facilitarán el estudio de los distintos elementos de una sola wiki.

5.2. Visualización

En WikiChron Monowiki se busca mostrar información compleja, como desgloses de un dato o distribuciones, y por tanto se requiere de un tipo gráfico más sofisticado y expresivo que la serie temporal.

A continuación, se explican los distintos tipos de gráficos incluidos en WikiChron Monowiki, y los cambios hechos en la aplicación inicial para incluir varias visualizaciones.

5.2.1. Nuevos gráficos utilizados

Los tres nuevos modos añadidos, además del gráfico de líneas, son los siguientes:

- **Líneas:** este modo es el que está implementado en WikiChron Classic. Se usa para representar aquellas métricas cuyo resultado es una serie temporal.
- **Barras:** es el tipo de gráfico más frecuente en WikiChron Monowiki. Se usa para representar aquellas métricas cuyo resultado es una lista de series temporales, siendo cada una de estas series, una categoría distinta que atiende a algún criterio. Como por ejemplo: usuarios que han editado una vez al mes y usuarios que han editado entre 2 y 4 veces al mes. Así, se tienen varias barras apiladas, una por cada categoría de la métrica en cuestión. A la hora de diseñar las métricas, hemos tenido en consideración que si las categorías de las métricas son excluyentes y ordenadas, los colores de las barras se representan en un solo color y aumentando la intensidad (en concreto, en una escala de verdes y azules) y si no lo son, se representan con colores diferentes y disjuntos. En definitiva, el uso de los colores pretende transmitir información de forma clara al usuario y que se entienda de un vistazo.

- **Mapas de calor:** se usan para representar distribuciones. Estas métricas devuelven una lista formada por los datos que generaran los ejes x, y, z y por el nombre que irá en la leyenda de la barra de color (eje z).
- **Series con áreas:** se usan en métricas que muestran porcentajes. Estas métricas devuelven los datos de la misma forma que lo explicado para los gráficos de barras. Este gráfico muestra varias series temporales coloreando el área existente entre una serie temporal y otra, de forma que se remarca la diferencia de alcance entre una serie temporal y otra.

5.2.2. Modificaciones para incluir los nuevos gráficos

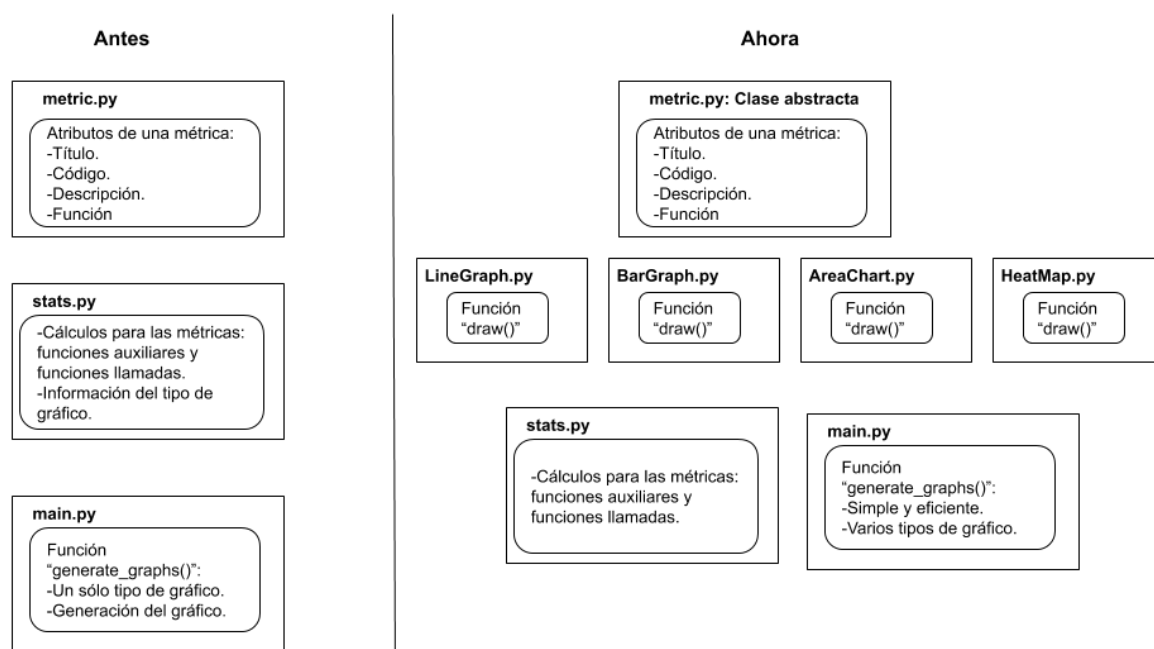


Figura 5.1: Generación de gráficos: antes y después de las modificaciones introducidas para la inclusión de distintos modos de visualización.

Para incluir estas visualizaciones, se ha modificado la estructura de la aplicación tal y como se muestra en la figura 5.1. Como se observa, en la aplicación original, existía un único tipo de métrica: `metric`, en el módulo `metric.py`, que incluía los distintos atributos de una métrica (título de la métrica, descripción, función en el módulo `stats.py` para calcularla y el identificador de la misma). Por otra parte, en esta versión anterior, el módulo `main.py`, incluía además de diversas funciones que hacen que la aplicación funcione, una función, llamada “`generate_graphs()`” que generaba la repre-

sentación gráfica de las métricas que era, por defecto, un gráfico de líneas para mostrar las series temporales de WikiChron Classic.

La nueva implementación se basa en la herencia de clases, de forma que ahora la clase `Metric` (en el módulo `metric.py`) es una clase abstracta, de la que heredan las cuatro subclases que representan los cuatro modos de visualización disponibles: `LineGraph`, `AreaChart`, `BarGraph` y `HeatMap`, todas ellas implementadas en módulos con el mismo nombre. Cada una de estas clases posee un método `draw()`, para poder generar un gráfico distinto en función del modo de visualización deseado. De esta forma, la función `generate_graphs()` en el módulo `main.py` tan sólo tiene que llamar al método `draw()` de la métrica en cuestión, aislando la lógica correspondiente a la generación de gráficos en el módulo correspondiente a cada tipo de visualización.

El fin último de esta implementación es estructurar el código de forma más clara, y aislar la lógica de generación de gráficos visuales de los módulos `stats.py` y `main.py`, además de permitir que el añadir formas de visualización nuevas sea escalable y eficiente.

Además, por la introducción de estos nuevos tipos de gráficos, nos hemos visto obligados a modificar otros aspectos, como la forma de realizar la descarga de los datos de los gráficos desde la aplicación. Ahora, todos los datos descargados son de una sola wiki, por lo que el nombre del archivo comprimido obtenido es igual a la wiki seleccionada y, dentro de este, tenemos un archivo por cada métrica elegida. Según el tipo de gráfico guardamos lo siguiente:

- Gráficos de líneas: la serie temporal (fecha en meses y su valor correspondiente).
- Gráficos de barras y los de áreas: la serie temporal de cada submétrica (fecha en meses y sus respectivos valores, uno por cada submétrica)
- Mapas de calor: la fecha en meses representando las filas, el valor del eje Y representando las columnas y su valor correspondiente, que sería el cálculo del eje Z

5.3. Métricas centradas en la distribución de usuarios registrados activos

Como se expone en el marco teórico, existen numerosos estudios, como el de Kittur et al. (2006), en los que se estudian uno o varios fenómenos de proyectos colaborativos atendiendo a tipos de usuarios (en el caso del artículo mencionado, se estudia el nivel de participación de los usuarios en función del número de ediciones del usuario).

Ya que los tipos de usuario establecidos en una comunidad colaborativa (administradores, usuarios nuevos, etcétera), están definidos siguiendo criterios que son propios de la misma (por ejemplo, en Wikipedia se elige a los administradores siguiendo un proceso de selección particularmente suyo, mencionado en el marco teórico), para este trabajo hemos decidido establecer tipos de usuario que crean una división basada en datos, y no dependiente de las normas de la comunidad.

Por tanto, una de las finalidades que tiene WikiChron Monowiki es permitir desgranar a los usuarios activos en un mes en varios tipos, atendiendo a diversos aspectos que pueden obtenerse a partir del análisis de los datos obtenidos del dump de Wikia.

Así, hemos implementado métricas en WikiChron Monowiki, que se encuentran en el menú de “Distribución de usuarios registrados activos” o “Distribution of active registered user” y permiten detectar diferentes tipos de editores atendiendo a los siguientes aspectos:

- La antigüedad del usuario en la wiki, lo que nos permite determinar si un usuario es nuevo o ha transcurrido cierto tiempo desde su llegada a la wiki.
- El tiempo transcurrido desde la última edición del usuario en la wiki, lo que permite distinguir entre usuarios que han vuelto a la wiki después de un período más corto o más largo de inactividad.
- El número de meses seguidos que lleva editando, lo que permite distinguir entre usuarios que editan más o menos frecuentemente.
- El tipo de página editado, lo que permite desgranar a los usuarios activos en un mes en función de si han editado en una página u otra. Es importante tener en cuenta que, las categorías no son excluyentes, como en el caso de las anteriores métricas, porque los usuarios pueden editar varios tipos de página al mes, por lo que en el caso de esta métrica, las categorías no son excluyentes.

Para todas estas métricas, nos quedamos solo con los usuarios registrados en la wiki y la suma de los usuarios de todas las categorías suman el total de usuarios activos registrados en cada mes, es decir, se desglosan los usuarios activos registrados mensuales en categorías diferentes según el propósito de cada métrica. Por esto, el gráfico más apropiado es el de barras, donde cada una de las barras de un mes representa a una categoría de usuario.

Además, hemos incluido una versión de estas métricas en porcentaje, que permite observar con más facilidad, cuál de todas las categorías de usuarios para cada métrica que mide la distribución de usuarios registrados activos, es más numerosa sobre el total de usuarios registrados activos cada mes.

A continuación, se proporciona una explicación detallada de la implementación de dichas métricas, con ejemplos de visualización para las mismas, en

la wiki de Cocktails ¹ de Fandom.

5.3.1. Usuarios activos según experiencia editando

La métrica de Usuarios activos según su experiencia editando o "Dist. of active registered users by editing experience", desgrana a los usuarios activos en un mes en cinco categorías, en función del número de ediciones que han hecho hasta el mes anterior (incluido el mismo):

- Usuarios activos muy experimentados: aquellos que han contribuido en un número de ediciones mayor o igual que 100.
- Usuarios experimentados: aquellos que han contribuido en al menos 25 ediciones y menos de 100.
- Usuarios activos maduros: aquellos que han contribuido en más de 4 y menos de 25 ediciones.
- Usuarios activos novatos: aquellos que han contribuido en menos de 5 ediciones.
- Usuarios nuevos: aquellos que contribuyen por primera vez en la wiki durante el mes en cuestión.

5.3.1.1. Implementación

Para implementar esta métrica, se calcula el número total de ediciones que ha hecho cada usuario en cada mes, y se filtra atendiendo a una condición (que se genera haciendo una llamada a una función auxiliar “generate_condition_users_by_number_of_edits”, y que establece la división mencionada en la sección anterior, en función del número total de ediciones hechas hasta el mes anterior), que divide a los usuarios en las cinco categorías mencionadas anteriormente.

Un ejemplo de visualización con esta métrica es la de la figura 5.2, que emplea la wiki de Cocktails.

¹https://cocktails.fandom.com/wiki/Cocktails_Wiki

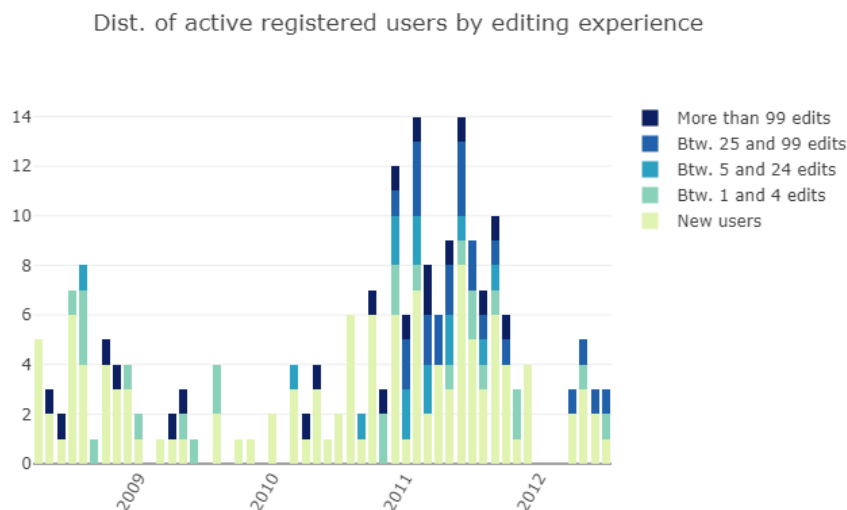


Figura 5.2: Métrica de distribución de usuarios activos registrados según su experiencia editando en la wiki de Cocktails, en el periodo temporal desde abril de 2008 hasta julio de 2012.

5.3.2. Usuarios activos por antigüedad

La métrica de “Dist. of active registered users by tenure”, también llamada Usuarios activos por antigüedad, agrupa a los usuarios activos de cada mes en función del momento en el que realizaron su primera edición en la wiki:

- Usuarios nuevos.
- Usuarios activos cuya primera edición fue entre hace 1 y 3 meses, ambos inclusivos.
- Usuarios activos cuya primera edición fue entre hace 4 y 6 meses, ambos inclusivos.
- Usuarios activos cuya primera edición fue entre hace 7 y 12 meses, ambos inclusivos.
- Usuarios activos cuya primera edición fue hace más de 12 meses.

5.3.2.1. Implementación

Para calcular cuántos usuarios activos pertenecen a cada categoría, se calcula cuántos meses han transcurrido desde el primer mes de actividad por

parte del usuario hasta cada mes en el que hay actividad registrada por parte de dicho usuario.

A continuación, se genera una condición basada en el número de meses transcurrido desde el primer mes de actividad, para detectar a qué categoría, de las anteriormente descritas, pertenece el usuario.

En la figura 5.3, se proporciona un ejemplo de visualización con esta métrica para la wiki de cocktails.

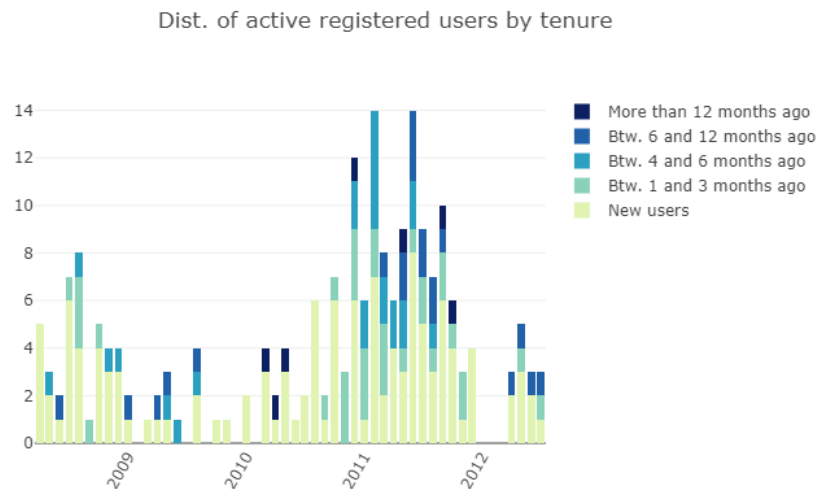


Figura 5.3: Métrica de distribución de usuarios activos registrados según su antigüedad en la wiki de Cocktails, desde abril de 2008 hasta julio de 2012.

5.3.3. Usuarios activos por retorno

La métrica de “Dist. of active registered users by date of the last edit”, o “Usuarios activos registrados que regresan”, categoriza a los usuarios activos en cada mes, en función del momento en el que hicieron su última edición en la wiki:

- Usuarios nuevos.
- Usuarios activos cuya última edición fue el mes anterior.
- Usuarios activos cuya última edición fue entre hace 2 y 3 meses, ambos inclusives.
- Usuarios activos cuya última edición fue entre hace 4 y 6 meses, ambos inclusives.
- Usuarios activos cuya última edición fue hace más de 6 meses.

5.3.3.1. Implementación

Para determinar a cuál de las categorías anteriores pertenecen los usuarios activos en cada mes, se calcula, para cada mes en el que hay registrada actividad del usuario, cuánto hace desde que editó por última vez. Para ello, se ordenan las fechas en las que contribuye cada usuario, de mayor a menor antigüedad.

Una vez obtenido el número de meses transcurridos entre un mes de actividad del usuario y el mes anterior al mismo en el que había actividad del usuario, se genera una condición basada en este número, con la que se determina qué usuarios activos en cada mes pertenecen a qué categoría de las definidas en la sección anterior.

En la figura 5.4, se muestra un ejemplo de visualización de esta métrica, empleando la wiki de Cocktails.

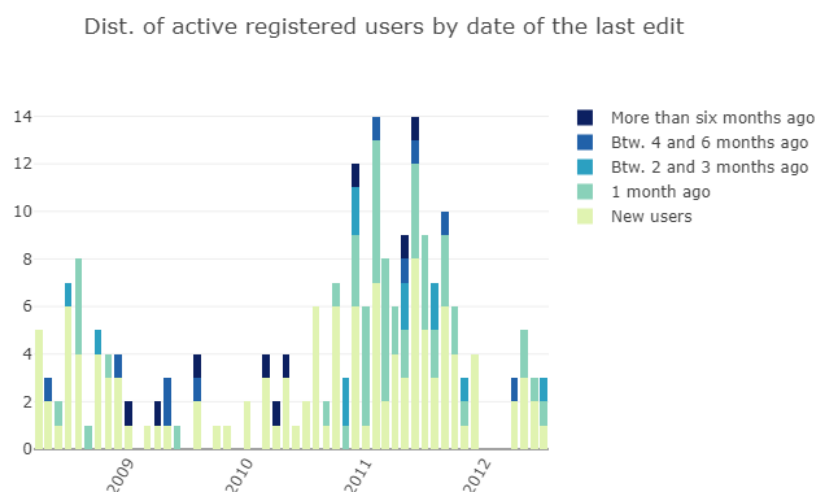


Figura 5.4: Métrica de distribución de usuarios activos registrados por retorno en la wiki de Cocktails, durante el período de abril de 2008 hasta julio de 2012.

5.3.4. Usuarios activos por racha de actividad

Esta métrica pretende visualizar el número de contribuidores que han editado un cierto número de meses consecutivos hasta el mes actual. Esto sirve para poder ver el compromiso de los usuarios dentro de la wiki. Tenemos una segunda versión de esta métrica que sería visualizar lo mismo pero solo con artículos.

Así, se definen, a continuación, las siguientes clases excluyentes, en función de en qué momento se lleva a cabo la última edición:

- Editores en el mes X cuya racha actual comienza en este mes, es decir, este mes ha editado, pero el anterior no.
- Editores en el mes X que han editado entre dos y tres meses seguidos, pero el cuarto no.
- Editores en el mes X que han editado entre 4 y 6 meses seguidos, pero el septimo no.
- Editores en el mes X que han editado mas de 6 meses seguidos.

5.3.4.1. Implementación

Para realizar la implementación de estas dos métricas hemos creado una función auxiliar llamada “current_streak_x_or_y_months_in_a_row”, donde lo primero que se hace es sumar al mes actual X e Y meses, en función de la categoría, y, desplazarlos X e Y posiciones y luego se comprueba si el mes actual está dentro de los meses que hemos desplazado y con eso conseguimos obtener a los usuarios que llevan editando entre X e Y meses seguidos. Esta función será usada usada también por alguna otra métrica relacionada explicada posteriormente.

En la figura 5.5, se muestra un ejemplo de visualización de esta métrica, empleando la wiki de Cocktails. Y en la figura 5.6, se muestra solo para artículos

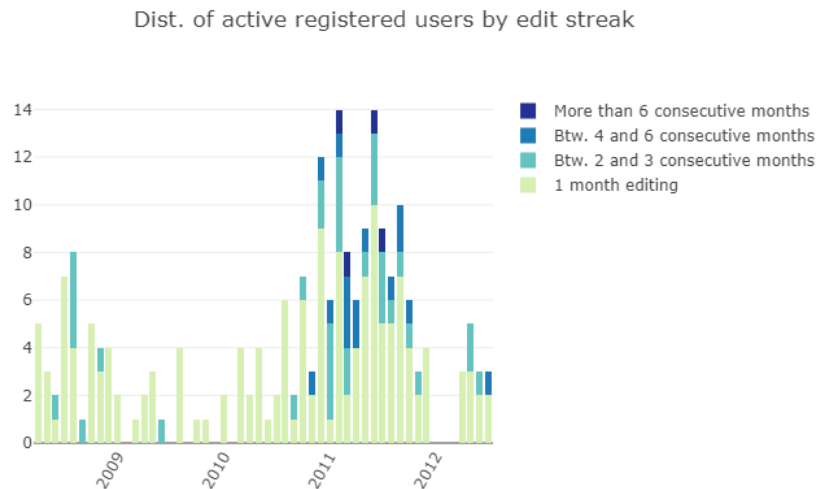


Figura 5.5: Métrica de distribución de usuarios activos registrados por racha de actividad, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.

5.3. Métricas centradas en la distribución de usuarios registrados activos 73

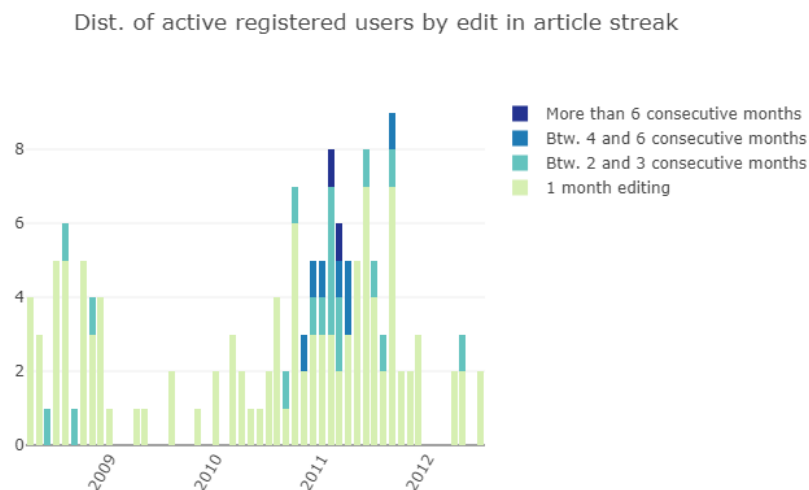


Figura 5.6: Métrica de distribución de usuarios activos registrados por racha de actividad en artículos, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.

Como hemos dicho anteriormente que estas métricas desglosan usuarios activos registrados, hemos calculado para todas ellas, excepto para la de usuarios activos por racha de actividad, su versión en porcentaje, es decir, el porcentaje que ocupa cada categoría de usuario dentro del total de usuarios activos registrados en cada mes.

Un ejemplo de visualización de unas de estas métricas sería la de la figura 5.7.

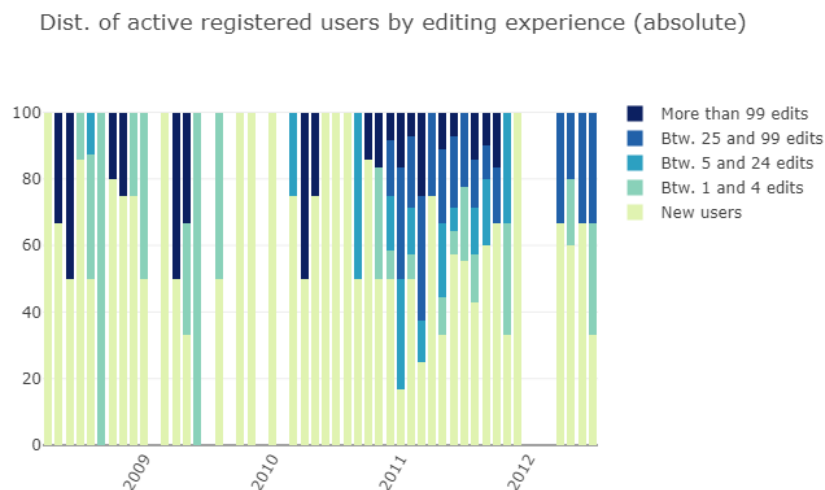


Figura 5.7: Métrica de distribución de usuarios activos registrados por su experiencia editando (versión absoluta) en la wiki de Cocktails, durante el período de abril de 2008 hasta julio de 2012.

5.3.5. Usuarios activos en función del tipo de página editada

Además de la motivación de distinguir entre qué tipos de páginas son los más editados por los usuarios activos en un mes, mencionada en la introducción a este tipo de métricas, investigando estudios existentes en comunidades colaborativas, encontramos el de Asta Zelenkauskaitė (2011), en el que los autores encuentran una relación entre la antigüedad de los usuarios y su predisposición para editar unos determinados tipos de páginas u otros.

En esta métrica nos centramos en los siguientes tipos de páginas, definidos en la sección de namespaces en las páginas de información de Fandom²:

- Article namespace: se identifica con el 0, y hace referencia al conjunto de páginas dedicadas a la escritura y edición de artículos.
- Article talk namespace: se identifica con el 1, y hace referencia al conjunto de páginas de discusión asociadas a un determinado artículo.
- User talk namespace: se identifica con el 3, y hace referencia al conjunto de páginas de discusión.
- User namespace: se identifica con el 2, y hace referencia al conjunto de páginas de usuario (páginas de perfil).

²<https://community.fandom.com/wiki/Help:Namespace>

- Template namespace: se identifica con el 10, y hace referencia al conjunto de páginas dedicadas a la escritura y edición de plantillas. Las plantillas, a su vez, son textos estándares cuya inclusión es necesaria en un determinado número de páginas.
- El resto de tipos de páginas, se incluyen dentro de otra categoría, llamada “otros”.

5.3.5.1. Implementación

Para computar el número de editores que contribuyen en páginas de artículos, discusión de artículos, discusión de usuarios, páginas de usuario, y páginas de plantilla, se emplea la misma dinámica, que se basa en la llamada a una función auxiliar, `filter_users_pageNS` (`data`, `index`, `page_ns`), que recibe, almacenado en el parámetro “`page_ns`”, el identificador del tipo de página. Posteriormente, esta función filtra el conjunto de datos original de Fandom (sometido al proceso de limpieza habitual), almacenado en la variable “`data`”, para que tan sólo incluya las ediciones realizadas en el tipo de página especificado por el parámetro “`page_ns`”. Así, devuelve una serie temporal que contiene, por una parte, el tiempo, y por otra, el número de editores que, en cada mes, edita una página de tipo “`page_ns`”.

Esta función es llamada por cinco funciones, una para cada categoría (`article`, `article talk`, `user`, `user talk` y `template`), cuyos pasos consisten, únicamente, en llamar a la función anterior con el parámetro correspondiente. Y con esto obtenemos una serie por cada categoría y el conjunto de ellas será devuelto en una lista.

En la figura 5.16, se muestra un ejemplo de visualización de esta métrica, en la wiki de Cocktails.

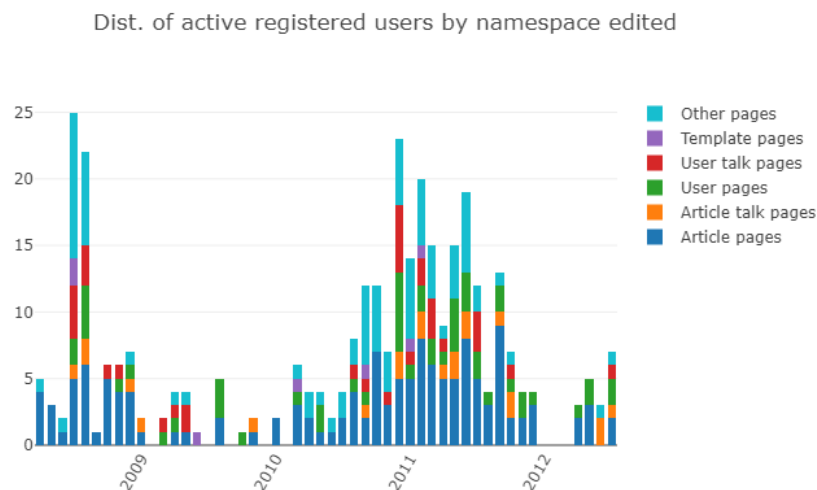


Figura 5.8: Métrica de usuarios activos registrados por ediciones según el tipo de página, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.

Nótese que, en esta métrica, si se suma el número de usuarios editando cada tipo de páginas, el total será mayor o igual que el número de usuarios contribuyendo, en al menos una edición, en ese mes. Y, a además, vemos como los colores de las barras no es monocromático con distintos grados de intensidad, ya que las categorías son independientes.

5.3.6. Ediciones en diferentes tipos de páginas

Con esta métrica podemos observar el número de ediciones que existen en cada tipo de página. Para representar esta funciones nos hemos decidido por un diagrama de barras apiladas, donde en el eje X tendríamos el tiempo, y en el eje Y (las barras). Para cada mes se apilarían distintas categorías con el número de ediciones para artículos, talk page de artículos, user Page, talk page de usuario, y agruparíamos el resto de tipo de páginas en otra categoría, rest. En algunas wikis hemos comprobado, que la categoría rest incluye muchas ediciones y como nos ayuda ver porque sucede esto, hemos extendido esta métrica en otra métrica que hemos nombrado (“By other namespace edited”), esta métrica en barras nos desglosa la categoría rest type, separandola en file, category, media wiki, template.

Nos parece interesante esta métrica, en relación con visualizar todas las ediciones que existen por meses, poder hacer una distinción de tipo de páginas que reciben mas ediciones en que meses o fenomenos curiosos como páginas que no se editan generalmente y en un mes reciben muchas ediciones.

5.3.6.1. Implementación

La implementación es igual a la descrita en la métrica del punto 5.3.5 pero cambiando el número correspondiente al tipo de página.

Un ejemplo de visualización para esta métrica sería la figura 5.9

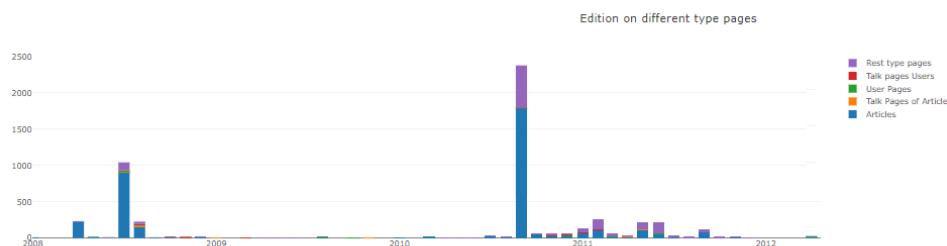


Figura 5.9: Métrica de usuarios activos registrados por ediciones según el resto de tipos de página, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.

5.4. Métricas centradas en la distribución de ediciones según tipos de usuarios registrados activos

En WikiChron Monowiki, hemos incluido un conjunto de métricas que permite discernir cuántas ediciones ha hecho en cada mes, cada clase de usuarios de las definidas en la sección anterior (Distribución de ediciones por experiencia editando, por antigüedad del usuario, por la racha de meses editando, por racha de meses editando en artículos, por la fecha de la última edición en la wiki y por tipo de página editado). Se encuentran en el menú de Distribución de ediciones según el tipo de usuario, o "Distribution of edits accross registered users".

De esta forma, WikiChron Monowiki permite visualizar de forma mensual, además de cómo de numeroso es cada tipo de usuario, en cuántas ediciones contribuyen estos tipos de usuarios activos. Con ello, se puede observar la cantidad de trabajo hecho por cada categoría de usuarios.

Cabe mencionar que estas métricas, al igual que las de la sección anterior, se han implementado teniendo en cuenta únicamente a los usuarios registrados.

La implementación seguida para computar este conjunto de métricas es idéntica, cambiando tan sólo en el cálculo de la información necesaria para determinar el tipo de usuario, que varía de una métrica de distribución de usuarios registrados activos a otra, y que es la que se ha explicado en la sección anterior.

El procedimiento seguido para calcular estas métricas de distribución de

ediciones según el tipo de usuario, consiste en calcular el número de ediciones mensual de cada usuario, para posteriormente filtrar el conjunto de datos de la wiki para mantener tan sólo la información de aquellos usuarios que pertenecen a una categoría u otra de cada métrica de distribución de usuarios registrados activos, de la forma explicada en cada una de las subsecciones anteriores, correspondientes a cada métrica de distribución de usuarios registrados activos.

Cuando se posee esta información, se agrupa por mes y se suma la cantidad de ediciones mensual por cada uno de los usuarios pertenecientes a dicha categoría. El proceso se repite para cada una de las categorías que tenga cada métrica de distribución de usuarios registrados activos.

También hemos incluido una versión de estas métricas en porcentaje, que calcula el porcentaje de ediciones que ha hecho cada categoría de usuario sobre el total de ediciones en cada mes, lo que facilita el ver qué tipo de usuario es más contribuyente.

A continuación, se muestran imágenes de estas métricas, empleando la wiki de Cocktails.

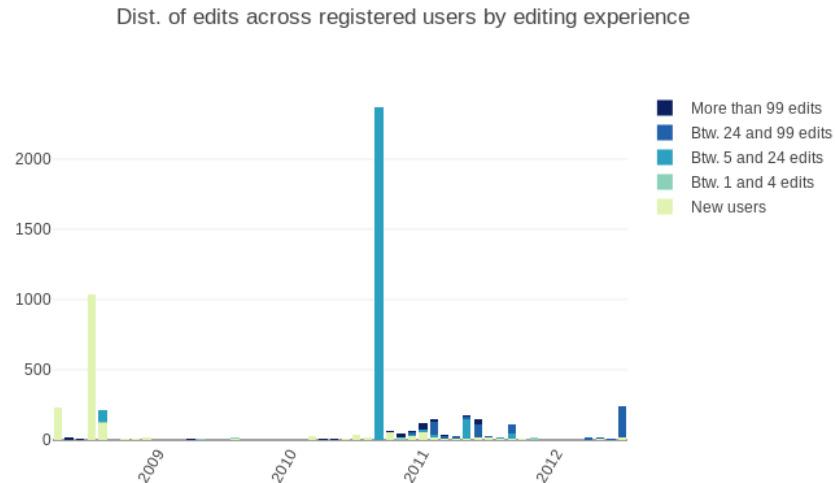


Figura 5.10: Métrica de distribución de ediciones por usuarios activos según su experiencia editando, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.

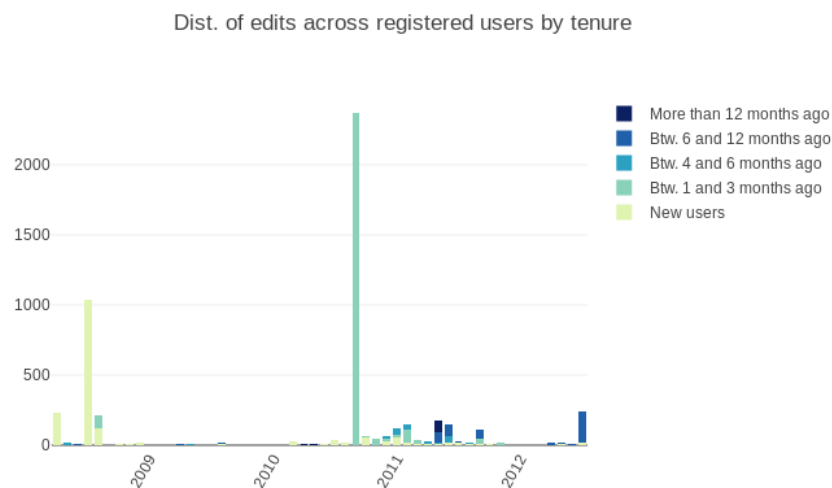


Figura 5.11: Métrica de distribución de ediciones por usuarios activos según su antigüedad, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.

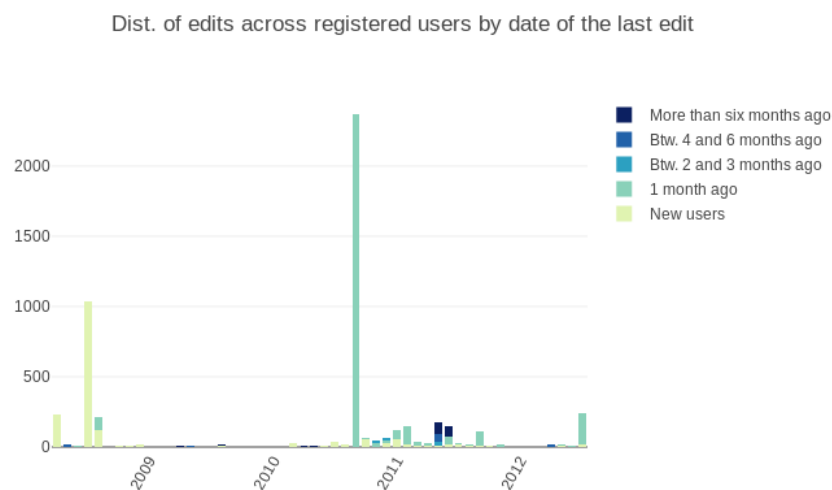


Figura 5.12: Métrica de distribución de ediciones por usuarios activos según su última edición, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.

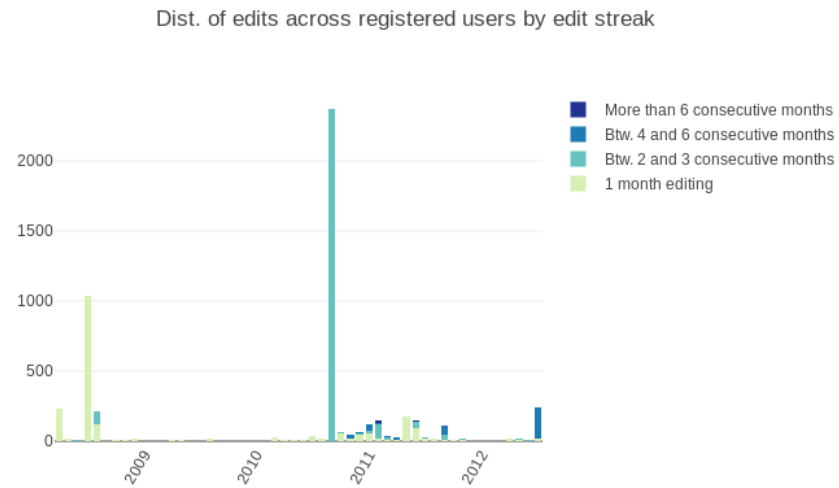


Figura 5.13: Métrica de distribución de ediciones por usuarios activos según su racha actual, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.

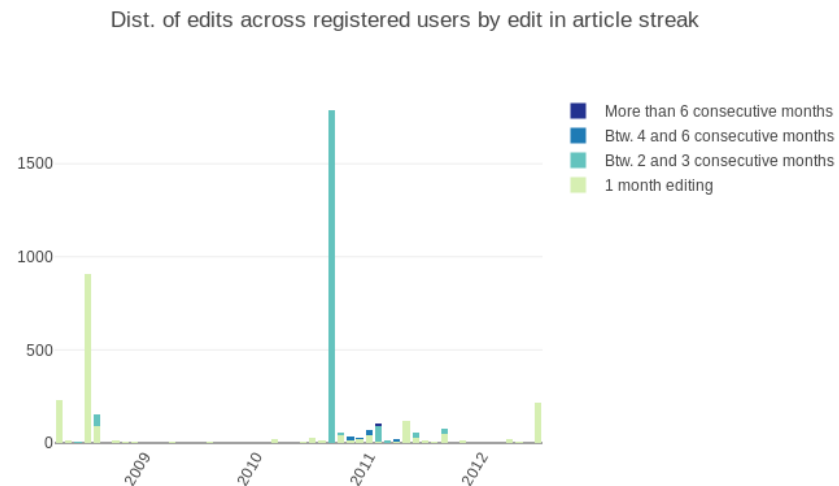


Figura 5.14: Métrica de distribución de ediciones por usuarios activos según su racha actual en artículos, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.

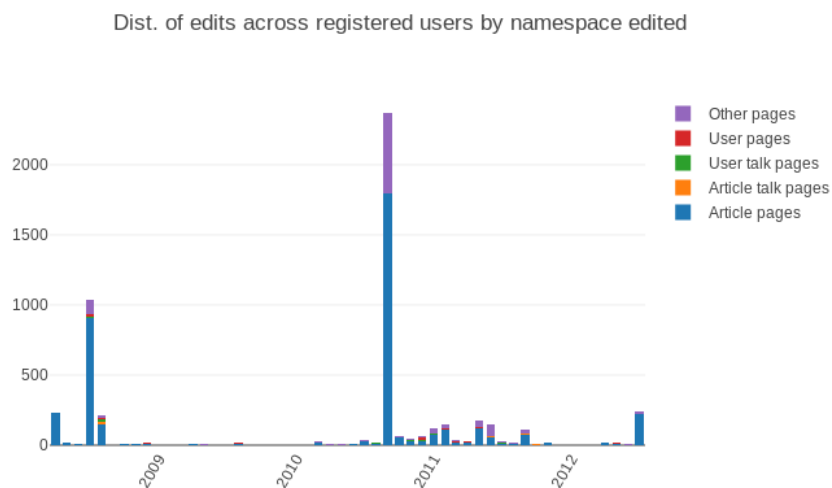


Figura 5.15: Métrica de distribución de ediciones por usuarios activos según el tipo de página editada, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.

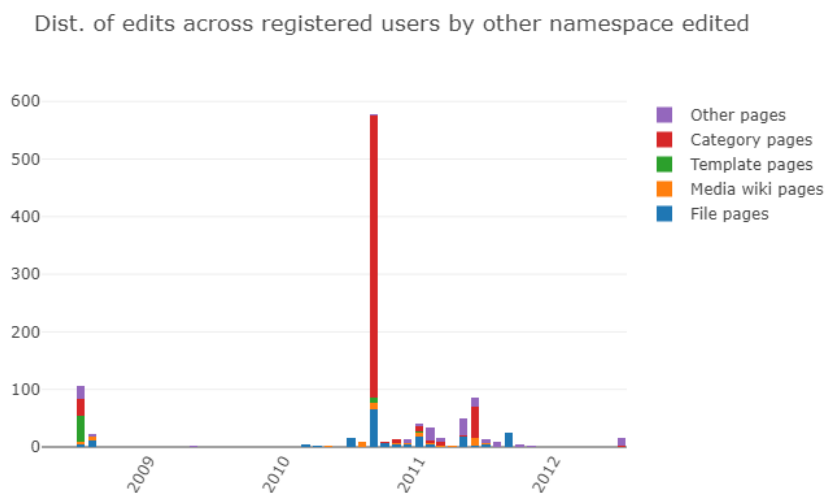


Figura 5.16: Métrica de distribución de ediciones por usuarios activos según el resto tipos de página editada, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.

5.5. Métricas centradas en las ediciones

A continuación, presentamos una serie de métricas centradas en el estudio de las ediciones y de la evolución de las páginas de WikiChron a lo largo del tiempo. Se distinguen de las anteriores en que éstas no miden la distribución de las ediciones según el tipo de usuario, sino en aspectos más genéricos. Estas métricas se encuentran en el menú de “Distribución de ediciones” o “Edits distribution”.

En concreto, es de interés en este trabajo de fin de grado, el poder visualizar la distribución de las páginas en función de cuántas veces se han revisado, para poder responder a preguntas como: ¿cuál es la media de revisiones que sufre una página? Cabe destacar que todas estas métricas se calculan en base a usuarios registrados.

5.5.1. Porcentaje de editores que han hecho un porcentaje fijado de ediciones

Tal y como se menciona anteriormente, en este trabajo de fin de grado, surge la necesidad de extrapolar a las wikis de Fandom, estudios ya realizados sobre otras comunidades colaborativas. En el análisis hecho en Yang et al. (2016), estudian, entre otros factores, la desigualdad de las contribuciones en Open Street Map ³, un proyecto colaborativo que tiene como objetivo, el crear un mapa del mundo de libre acceso y modificación. Para estudiar cómo evoluciona la desigualdad de participación en esta comunidad, emplean la medida de desigualdad estadística básica, el coeficiente de Gini. No obstante, para aportar información acerca de cómo es la estructura demográfica subyacente, como complemento al coeficiente de Gini, emplean otra métrica, que les permite observar qué porcentaje de contribuidores ha contribuido en la creación de un 50 %, 80 %, 90 % y 95 % de las ediciones. Así, hemos procedido a la creación de una métrica similar, que determina el porcentaje de contribuidores que contribuyen a la creación de un 50 %, 80 %, 90 % y 99 % de las ediciones de la wiki, para emplearla como complemento del coeficiente de Gini en análisis posteriores.

Hemos implementado dos versiones de esta métrica: La primera versión, determina el porcentaje de editores que contribuyen a la creación de cada uno de los porcentajes (50 %, 80 %, 90 % y 99 %) del total de ediciones hasta el momento en el tiempo de vida de la wiki. Es decir, para computar esta métrica, se emplea el número acumulado de ediciones que ha hecho cada contribuidor en cada mes. En la segunda versión de esta métrica, se emplea el número de ediciones que ha hecho cada usuario en cada mes, sin acumular, para determinar qué porcentaje de usuarios ha contribuido a la creación de un 50 %, 80 %, 90 % y 99 % de las ediciones totales del mes, sin acumular.

³<https://www.openstreetmap.org>

5.5.1.1. Implementación

Para ambas métricas, se sigue una dinámica similar. Para la métrica sobre el total de ediciones, se parte del conjunto de datos original de la wiki, y se obtiene el número de ediciones total que ha hecho cada usuario hasta cada mes en el que se registra actividad de dicho usuario. Luego, se obtiene el total de ediciones hechas hasta dicho mes, y se calcula el porcentaje de ediciones que ha hecho el usuario. Posteriormente, agrupando por mes, se ordena a los usuarios activos de cada mes de forma descendente de acuerdo con el porcentaje de ediciones que han hecho. Una vez ordenados, se calcula el acumulado de estos porcentajes en una columna aparte, y este valor se emplea para filtrar el conjunto de datos en función del número de usuarios necesarios para poder alcanzar el 50 %, 80 %, 90 % y 99 % de las ediciones.

Para la versión sobre el número mensual de ediciones, se procede de forma idéntica. Tan sólo varía el cómputo del número de ediciones hechas por el usuario, que es el número de ediciones que ha hecho el usuario en dicho mes, y no el acumulado hasta dicho mes. Además, para computar el porcentaje, se emplea el total de ediciones hechas en dicho mes, y no el acumulado.

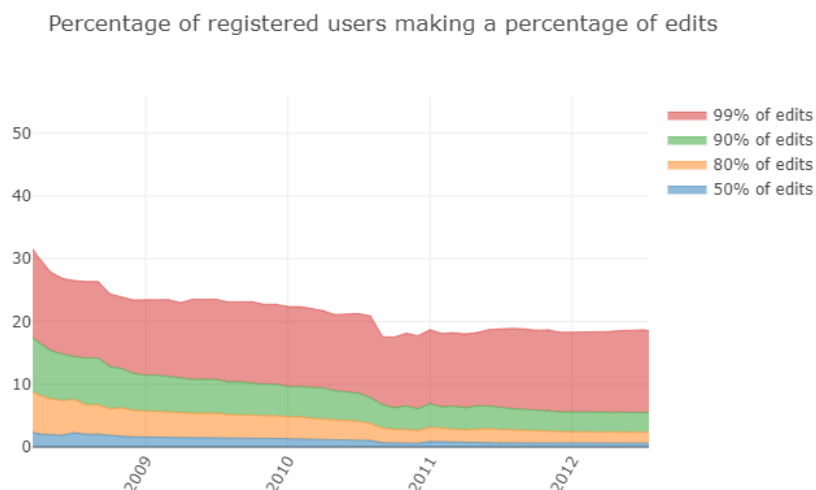


Figura 5.17: Métrica de porcentaje de editores por el porcentaje de ediciones (versión acumulada) en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.

Percentage of registered users making a percentage of edits (monthly)

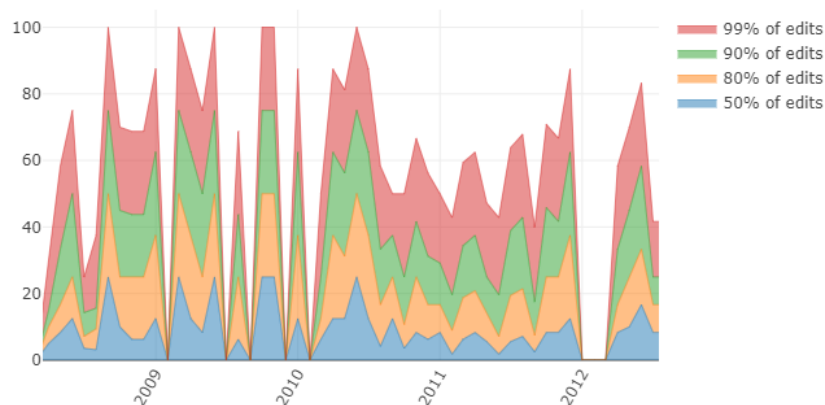


Figura 5.18: Métrica de porcentaje de editores por el porcentaje de ediciones (versión mensual) en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.

En la figura 5.17, se muestra un ejemplo de visualización con esta métrica en su versión acumulada, para la wiki de Cocktails, y en la figura 5.18, se muestra un ejemplo de visualización de la métrica en su versión mensual, para la wiki de Cocktails.

5.5.2. Distribución de ediciones entre usuarios registrados

Esta métrica lo que pretende visualizar es el número de editores que han hecho el mismo número de contribuciones. Para eso proponemos un heatmap (mapa de calor) donde el eje X sería el tiempo, el eje Y, el número de contribuciones y el eje Z, es decir, el color bar, sería el número de editores con el mismo número de contribuciones en un mes concreto. Con esta métrica podemos ver el nivel de participación de los contribuidores.

A a hora de visualizar esta métrica hemos observado que es complicado que varios usuarios realicen exactamente el mismo número de contribuciones y que a partir de un número de contribuciones realizadas, los usuarios que las realizaban, eran muy escasos y esto nos impedía visualizar correctamente esta métrica, por lo tanto, hemos optado por cortar el eje “Y” en 110 y en agrupar el número de las ediciones en intervalos de diez en diez, es decir, hemos sumado todos los usuarios que realizan un rango determinado de ediciones.

Un ejemplo de visualización sería el de la figura 5.19 para la wiki de Cocktails.

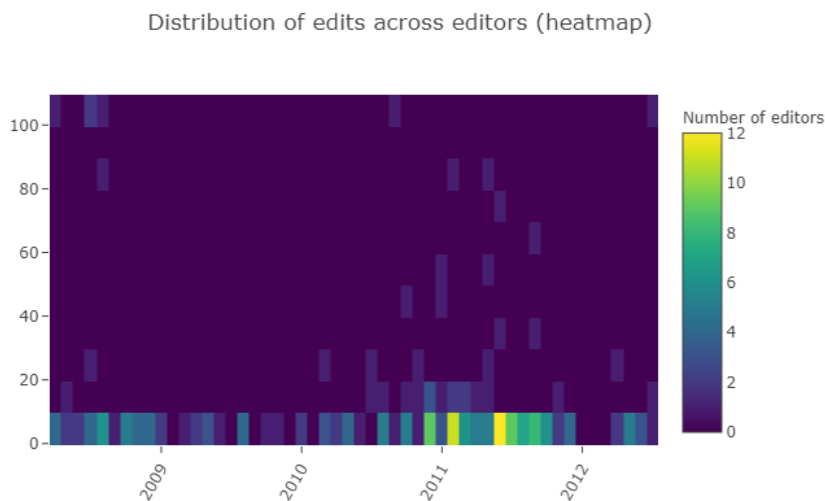


Figura 5.19: Métrica de distribución de usuarios entre ediciones, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.

5.5.3. Bytes añadidos y eliminados en los artículos

Estas métricas pretenden visualizar el número de artículos que añaden o eliminan los mismos bytes mensualmente. Para este caso, también proponemos un mapa de calor, donde el eje X sería el tiempo, el eje Y, el número de bytes añadidos o eliminados y el eje Z, el número de artículos que tienen esos bytes añadidos o eliminados.

La razón por la que hemos creado dos gráficas, una para los bytes añadidos y otra para los eliminados y no representarlo todo en una misma gráfica es porque el mapa de calor con un eje Y negativo no tiene una buena representación y por eso para la métrica de los bytes eliminados, se indica en positivo los bytes eliminados por mes.

5.5.3.1. Implementación

Lo primero ha sido calcular la diferencia de bytes entre las ediciones de un mismo artículo, a continuación, separamos, por un lado, la diferencia positiva incluyendo el cero, es decir, los bytes añadidos, y, por otro, la diferencia negativa, los bytes eliminados. Para los bytes eliminados, además calculamos su valor absoluto, para cambiarlos a números positivos.

Por la misma razón explicada en el punto 5.5.2, el límite lo ponemos en 1000, tanto para los bytes añadidos como para los eliminados y creamos intervalos de 100 en 100.

Un ejemplo de visualización para estas métricas son las figuras 5.20 y

5.21

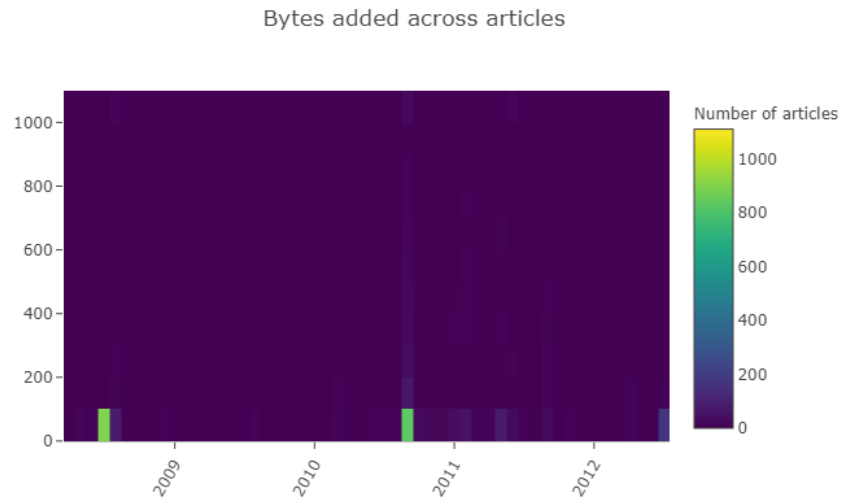


Figura 5.20: Métrica de bytes añadidos entre artículos, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.

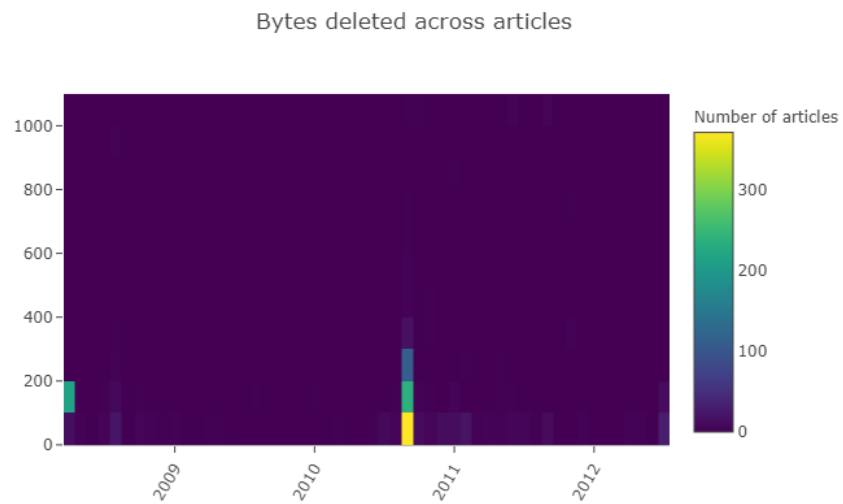


Figura 5.21: Métrica de bytes eliminados entre artículos, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.

5.5.4. Distribución de usuarios registrados según el número de artículos editados mensualmente.

Esta métrica permite visualizar cuánta gente edita el mismo número de artículos al mes, para representarla hemos utilizado un heatmap donde en el eje X, tendríamos el tiempo, en el eje Y el número de editores que han editado un mismo artículo, y en el eje Z indicaría el número de artículos que fueron editados por Y editores en el mes X. Esta función nos resulta útil para ver los usuarios al mes que editan el mismo número de artículos y como con el paso del tiempo, tiende a disminuir por un lado el número de artículos editados, o los usuarios que inicialmente coincidían editando un número de artículos elevado, cada vez coinciden menos con otros usuarios porque reducen sus ediciones.

5.5.4.1. Implementación

Lo primero que hacemos es contar el número de usuarios que hay por artículo y luego obtenemos el número de artículos en los que han editado el mismo número de usuarios.

Un ejemplo de visualización de esta métrica sería la figura 5.22

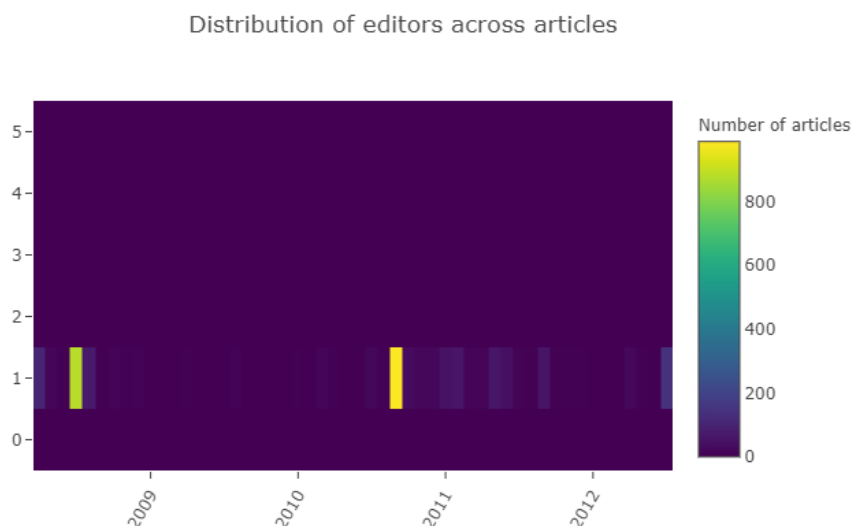


Figura 5.22: Métrica de distribución de usuarios registrados entre artículos por mes, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.

5.5.5. Distribución de las ediciones en páginas

Esta métrica nos permite observar, para todas las páginas, las ediciones que existentes a lo largo del tiempo, para eso hemos decidido implementarlo como un heatmap, en el que en el eje X tendríamos el tiempo, en el eje Y las ediciones, y en el eje Z tendríamos las páginas que han sido editadas Y veces en el mes X. Para mejorar la visualización de esta métrica hemos implementado el eje Y con bins, es decir, hemos creado intervalos, juntando ediciones, y para su valor de Z sumando todas las paginas que han sido editadas en el intervalo.

5.5.5.1. Implementación

Lo primero es obtener el número de ediciones en cada página, y, a continuación, calculamos cuántas páginas coinciden con cada intervalo de ediciones.

Un ejemplo de visualización para esta métrica sería la figura 5.23

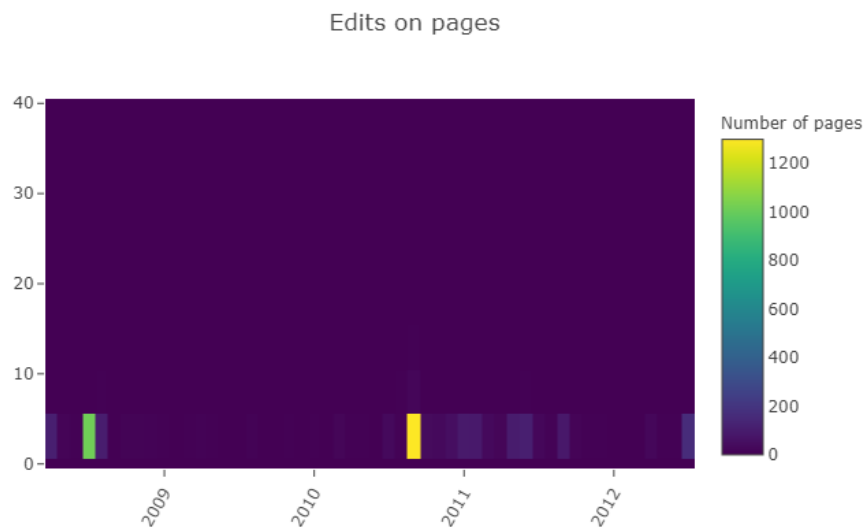


Figura 5.23: Métrica de ediciones en páginas, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.

5.5.6. Distribución de revisiones en páginas

Con esta métrica podemos visualizar, para todas las páginas, las revisiones ⁴ efectuadas a lo largo del tiempo, hemos implementado esta métrica

⁴Cualquier edición dentro de una página, excepto la primera.

mediante un heatmap en el que en el eje X tendríamos el tiempo, en el eje Y las revisiones, y en el eje Z tendríamos las páginas que han sido revisadas Y veces en el mes X. Esta métrica igual que la anterior la hemos implementado de tal manera que el eje Y sean intervalos, pero en vez de agrupando ediciones, en este caso lo hemos hecho con revisiones. Esta métrica nos parece útil verla en contraposición de la anterior para visualizar páginas que han sido creadas y nunca más revisadas o páginas que tienen muchas revisiones por detrás y puede ser un indicio de fiabilidad respecto a otras páginas.

La única diferencia que hay con la métrica explicada en el punto 5.5.5 es que quitamos la primera edición realizada en cada una de las páginas, es decir, la creación de la página.

Un ejemplo de visualización de esta métrica sería la figura 5.41

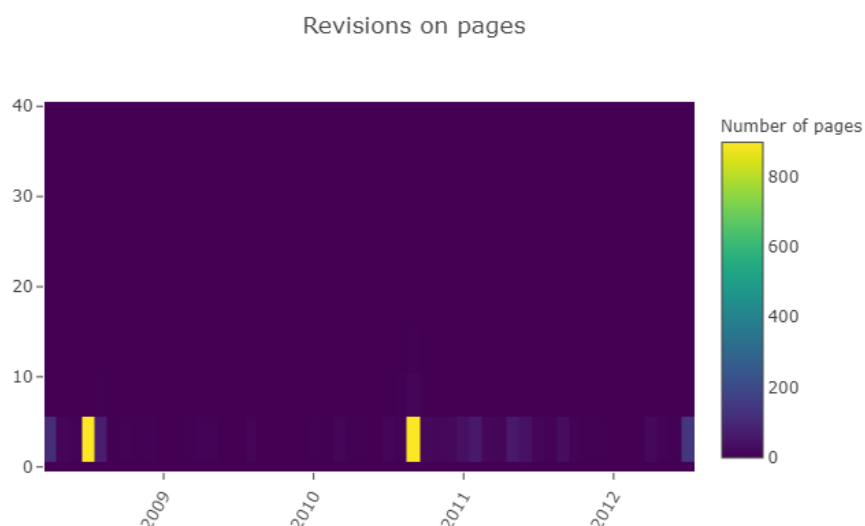


Figura 5.24: Métrica de revisiones en páginas, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.

Para hacer los intervalos mencionados en los mapas de calor descritos, hemos realizado:

1. Lo primero de todo hemos creado una lista que va desde el mínimo del eje Y hasta el máximo por intervalos, por ejemplo, si queremos que los intervalos vayan de 10 en 10, pues creamos una lista que sea $[0, 10, 20, \dots]$.
2. Con esta lista determinamos, el rango al que pertenece cada cálculo obtenido según la métrica a calcular.

3. Luego, agrupamos todos los cálculos que pertenezcan al mismo rango por cada mes y obtenemos el total.

Una vez que tenemos la serie descrita anteriormente, para calcular el eje Z, iteramos esa serie, obteniendo el valor máximo y mínimo de cada rango agrupado y generando una matriz, que, por cada mes rellene todas las posiciones entre el mínimo y el máximo con el mismo número que tiene el rango. Luego trasponemos esa matriz para que por cada valor de Y tengamos el valor de Z en todas las X. El eje X es el index pasado como parámetro a la métrica y la Y es una lista entre el mínimo y el máximo de todos los valores de la columna que queremos representar en el eje Y.

5.6. Métricas centradas en unidades de conocimiento

A continuación, presentamos una serie de métricas centradas en el estudio del conocimiento en una wiki. Tal y como se menciona en el marco teórico, existen diversos estudios sobre la evolución del conocimiento en proyectos muy exitosos, como Wikipedia. En el trabajo realizado por (Chhabra y Iyengar, 2018), en el que se define un factoides como una unidad de conocimiento en sí misma. Estos autores emplean los diferentes factoides de las 100 páginas más populares de Wikipedia para estudiar su evolución, y cómo de relacionada está la presencia de un factoides determinado en las primeras revisiones de la página con la aparición de un nuevo factoides en revisiones posteriores de la misma.

Estos autores determinan que un enlace a otra página es un factoides, y que un nombre propio también lo es, ya que ambos pueden definirse como unidades de conocimiento. En WikiChron Monowiki, hemos adoptado esta última parte de la definición, por lo que consideramos cada nombre propio como un factoides.

La inclusión de métricas para el estudio del conocimiento en WikiChron, tiene como objetivo el poder visualizar la evolución temporal en el número de factoides existentes en la wiki, y qué grupos de usuarios aportan o borran el conocimiento existente en los artículos de la wiki.

Además, también hemos incluido métricas que permiten visualizar la evolución temporal en el número de enlaces e imágenes en los artículos de una wiki, ya que creemos que un artículo bien dotado de estos dos elementos es un artículo más elaborado y rico que uno con menos enlaces e imágenes.

Para llevarlas a cabo, se necesita acceder al contenido de las ediciones del dump de Wikia correspondiente a la wiki, esto está explicado en el punto 5.6.1

Estas métricas se encuentran disponibles en el menú de Distribución de factoides según el tipo de usuario o "Distribution of factoids across registered users", en el que se incluyen diferentes métricas para el estudio de la

distribución de factoides por tipo de usuario en forma de gráfico de barras, y en el menú de Factoides o "Factoids", en el que se proporcionan métricas para el estudio de la evolución temporal en el número de factoides añadidos y borrados en la wiki, en forma de gráfico de líneas. Las métricas sobre s e imágenes se encuentran en el menú de Contenido de ediciones o "Edit content".

En las secciones a continuación, se explica en primer lugar en qué han consistido las modificaciones en el parser de WikiChron para poder acceder al contenido de las ediciones del dump de Wikia. En último lugar, se explica en qué consisten las métricas proporcionadas para el estudio del conocimiento.

5.6.1. Acceso al contenido de las ediciones del dump de Wikia

Para realizar este estudio, hemos tenido que modificar el parser "wiki_dump_parser" ⁵ que genera el csv con los datos de las wikis. Lo primero que hemos hecho ha sido obtener el texto en bruto del archivo XML de cada una de las revisiones realizadas. A continuación, lo hemos limpiado, quitándole el marcado y luego hemos clasificado las palabras para así obtener los nombres propios, que serán tratados como factoides. Para realizar esto último, hemos usado una librería de python centrada en el procesamiento del lenguaje natural, como es NLTK. Esto está probado tanto para inglés como para español.

Finalmente, el csv generado por el nuevo parser tendría dos columnas más, una de ellas con el texto en bruto (sin limpiar), y otra columna con los factoides.

5.6.2. Conteo de factoides añadidos y borrados

Este menú de WikiChron Monowiki presenta cuatro métricas centradas en el estudio de la evolución temporal de los factoides. Es importante mencionar que se centran en el espacio de nombres principal de la wiki, es decir, únicamente se estudia el conocimiento presente en los artículos, ya que son aquellas páginas donde por norma general se registra el conocimiento de una wiki. Las métricas disponibles en este menú son:

⁵https://github.com/laudf/wiki-scripts/blob/master/wiki_dump_parser/wiki_dump_parser.py

5.6.2.1. Factoides añadidos y Total de factoides añadidos

contributor_id	page_id	factoids	added_factoids
33456	1	{Cat, Dog, Duck, Seal}	{Cat, Dog, Duck, Seal}
5430	1	{Cat, Duck, Bear}	{Bear}
2	1	{Cat, Dog, Duck, Seal}	{Dog, Seal}
2	2	{Hot, Cold}	{Hot, Cold}
5430	3	{Orange, Green}	{Orange, Green}
1	3	{Blue, Red}	{Blue, Red}
6712	4	{Apple, Lemon, Banana}	{Apple, Lemon, Banana}
33456	4	{Apple, Lemon}	{}
2	4	{Lemon, Orange}	{Orange}
1	4	{Strawberry, Watermelon}	{Strawberry, Watermelon}

Figura 5.25: Ejemplo gráfico de factoides añadidos en distintas revisiones de varias páginas. Nótese que de una revisión a otra, pueden añadirse y borrarse factoides

Estas métricas muestran el número mensual y el acumulado de factoides añadidos. Para calcular el número de factoides añadidos, tal y como se muestra en la figura 5.25, sobre el csv original de la wiki (que contiene una fila con la información perteneciente a cada edición), se agrupan las páginas por su identificador, que es único. Una vez agrupadas, se cuenta cuántos factoides han sido añadidos en ediciones consecutivas de una misma página y se añade el cálculo a nueva columna del DataFrame de Pandas. Finalmente, se agrupa por mes y se calcula el número mensual de factoides añadidos para la métrica de Added factoids y el número total de factoides añadidos para la métrica de Total added factoids. Las figuras 5.26 y 5.27 muestran un ejemplo de estas métricas.

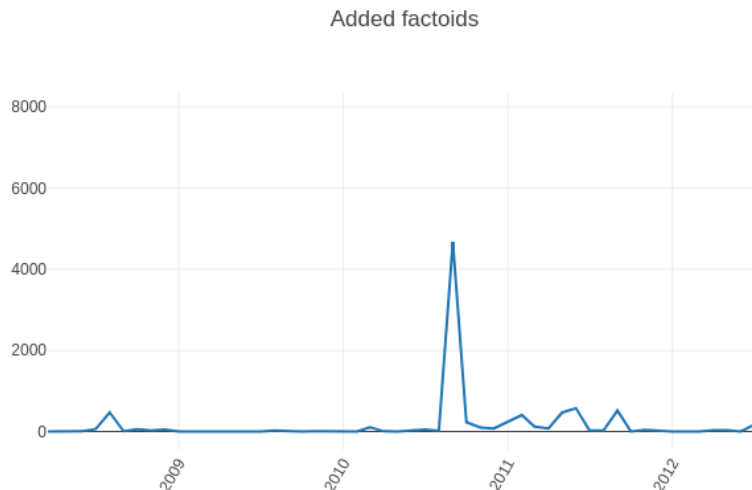


Figura 5.26: Métrica de Added factoids en la wiki de Cocktails, en el período temporal de abril de 2008 a julio de 2012.

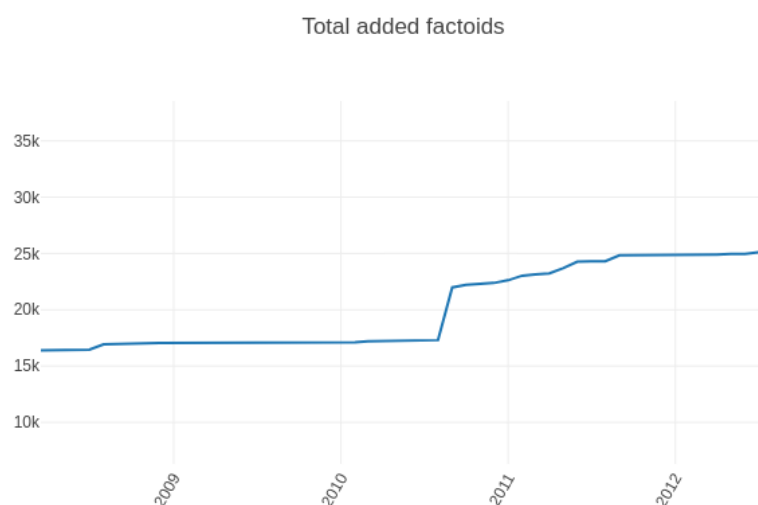


Figura 5.27: Métrica de Total added factoids en la wiki de Cocktails, en el período temporal de abril de 2008 a julio de 2012.

5.6.2.2. Factoides borrados y Total de factoides borrados

contributor_id	page_id	factoids	deleted_factoids
33456	1	{Cat, Dog, Duck, Seal}	{}
5430	1	{Cat, Duck, Bear}	{Dog}
2	1	{Cat, Dog, Duck, Seal}	{Bear}
2	2	{Hot, Cold}	{}
5430	3	{Orange, Green}	{}
1	3	{Blue, Red}	{Orange, Green}
6712	4	{Apple, Lemon, Banana}	{}
33456	4	{Apple, Lemon}	{Banana}
2	4	{Lemon, Orange}	{Apple}
1	4	{Strawberry, Watermelon}	{Lemon, Orange}

Figura 5.28: Ejemplo gráfico de factoides borrados en distintas revisiones de una misma página.

Estas métricas muestran el número mensual y total de factoides borrados. Para calcularlo, de forma análoga a las métricas de Added factoids y Total added factoids (ver figura 5.28), se agrupa por el identificador de la página, y se determinan cuántos factoides se han borrado en ediciones consecutivas de una misma página. Teniendo esta información, se agrupa por mes y se calcula el número mensual y total de factoides borrados. Las figuras 5.29 y 5.30 muestran un ejemplo con estas métricas.

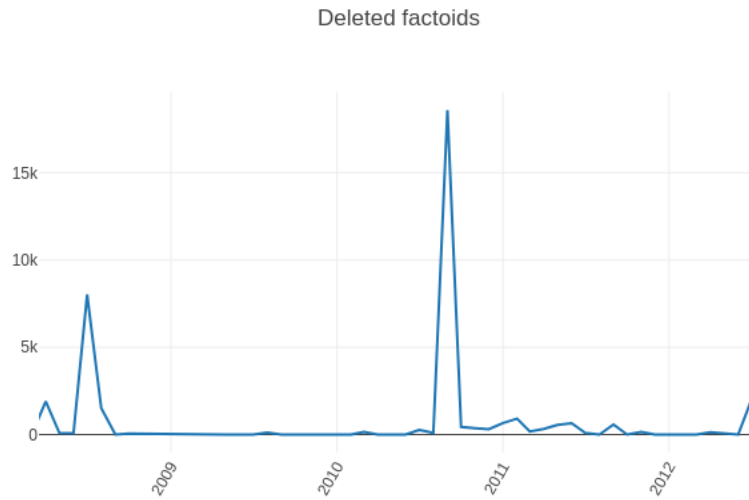


Figura 5.29: Métrica de Deleted factoids en la wiki de Cocktails en el período temporal de abril de 2008 a julio de 2012.

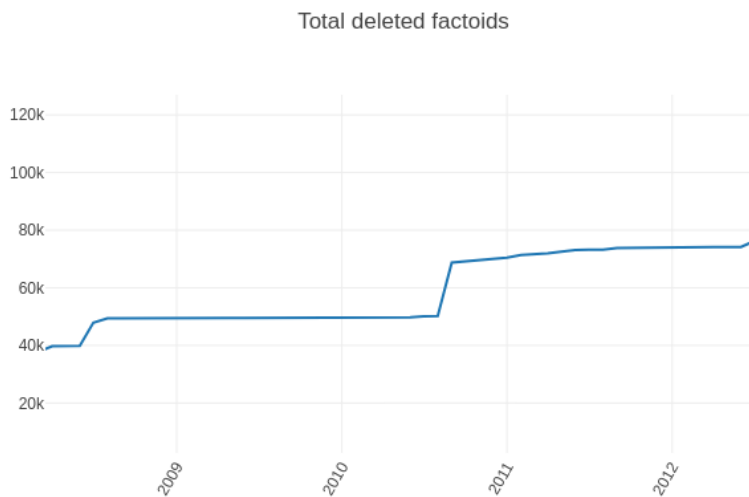


Figura 5.30: Métrica de Total deleted factoids en la wiki de Cocktails en el período temporal de abril de 2008 a julio de 2012.

5.6.3. Distribución de factoides añadidos y borrados por tipos de usuario

Este menú incluye ocho métricas que permiten estudiar cómo se distribuyen los factoides según el tipo de usuario que los añade o elimina.

5.6.3.1. Distribución de factoides añadidos por tipos de usuario

En esta categoría se encuentran las siguientes métricas:

- Nótese que esta métrica no está computada para el tipo de usuario `Users by namespace edited`, dado que las métricas en factoides se centran en el espacio de nombres de los artículos. Por esta misma razón, tampoco se computa para el tipo de usuario `Users by article edit streak`. Las figuras 5.31, 5.32, 5.33 y 5.38 muestran un ejemplo de uso de estas métricas.

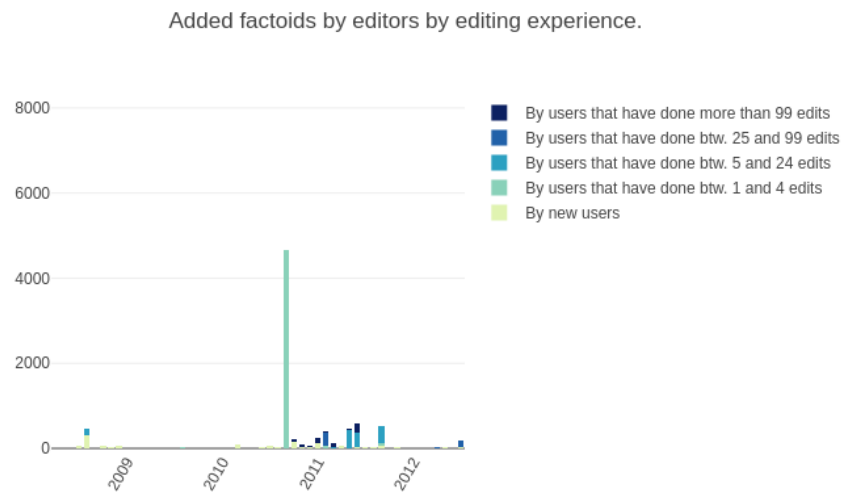


Figura 5.31: Métrica de Added factoids by active users by editing experience en la wiki de Cocktails en el período temporal de abril de 2008 a julio de 2012.

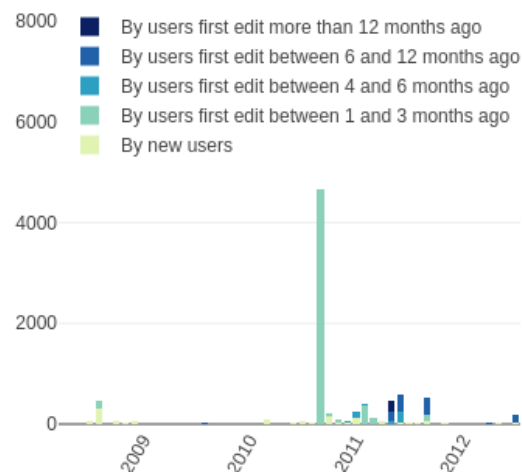


Figura 5.32: Métrica de Added factoids by active users by tenure en la wiki de Cocktails en el período temporal de abril de 2008 a julio de 2012.

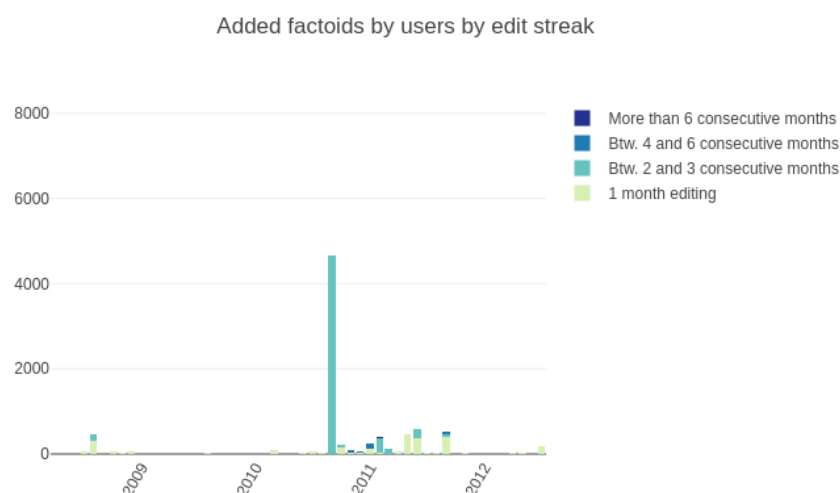


Figura 5.33: Métrica de Added factoids by users by edit streak en la wiki de Cocktails en el período temporal de abril de 2008 a julio de 2012.

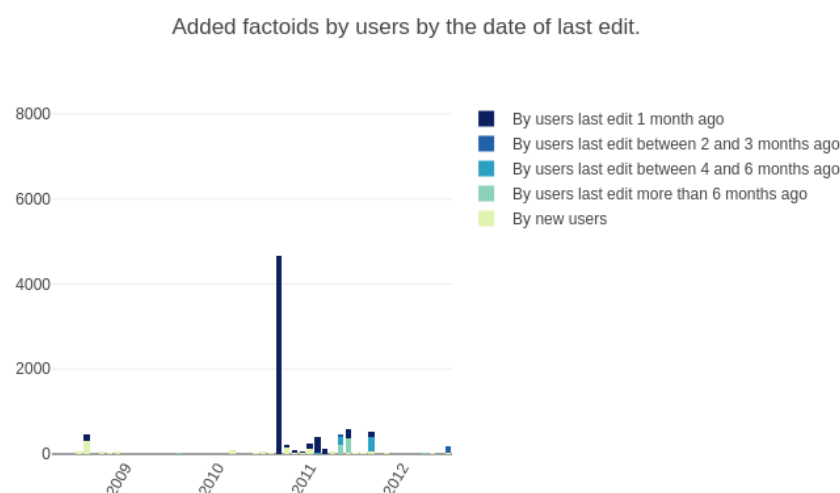


Figura 5.34: Métrica de Added factoids by active users by the date of the last edit en la wiki de Cocktails en el período temporal de abril de 2008 a julio de 2012.

5.6.3.2. Distribución de factoides borrados por tipos de usuario

En esta categoría se encuentran las métricas de Deleted factoids by editor experience, Deleted factoids by users by tenure, Deleted factoids by users by edit streak y Deleted factoids by users by the date of the last edit. Al igual que ocurre con las métricas que estudian la distribución de factoides añadidos por tipos de usuario, en este tipo de métricas se comienza de la misma forma que en las métricas de Deleted factoids y Total deleted factoids, explicadas

en la sección anterior (ver figura 5.28). Partiendo del csv original de la wiki (eliminando usuarios anónimos y páginas que no son artículos), se comienza agrupando por página, y se computa el número de factoides borrados de una edición a la siguiente sobre una misma página. Posteriormente, se filtra este DataFrame de Pandas para mantener tan sólo la información acerca de aquellos usuarios que cumplen la condición necesaria para pertenecer a una categoría u otra del tipo de usuario para el que se quiere observar la distribución de factoides borrados. Teniendo este DataFrame filtrado, se agrupa por mes y se suma el número total de factoides borrados por cada categoría de dicho tipo de usuario.



Figura 5.35: Métrica de Deleted factoids by active users by editing experience en la wiki de Cocktails en el período temporal de abril de 2008 a julio de 2012.

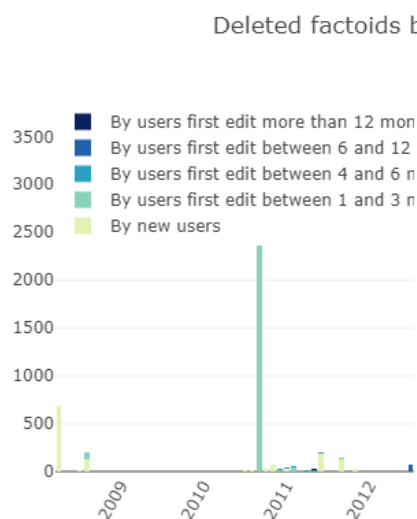


Figura 5.36: Métrica de Deleted factoids by active users by tenure en la wiki de Cocktails en el período temporal de abril de 2008 a julio de 2012.

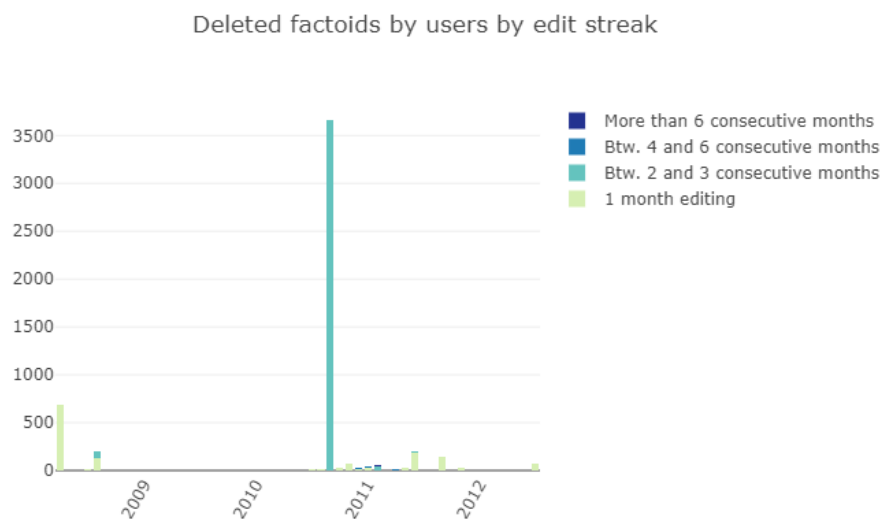


Figura 5.37: Métrica de Deleted factoids by active users by edit streak en la wiki de Cocktails en el período temporal de abril de 2008 a julio de 2012.

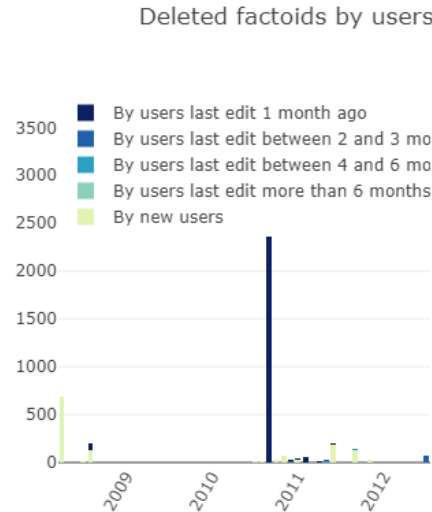


Figura 5.38: Métrica de Deleted factoids by active users by the date of the last edit en la wiki de Cocktails en el período temporal de abril de 2008 a julio de 2012.

5.7. Métricas centradas en enlaces e imágenes en artículos

Tras añadir las dos nuevas columnas en el csv como hemos explicado en el punto 5.6.1 (Edit_content y Factoids), hemos usado esta primera que contiene el contenido de cada edición para generar tres nuevas métricas, filtrando todo el texto de la edición obtenemos los enlaces y las imágenes, esto antes no se podía realizar debido a que en los csv no se guardaba el texto completo.

La primera métrica contabiliza los enlaces que hay en cada página, tanto internos, que son los que nos dirigen a la misma o otra página de la wiki y se detectan porque en la edición aparece representado entre doble corchete y los externos, que nos dirigen a una página externa y se representan mediante una url.

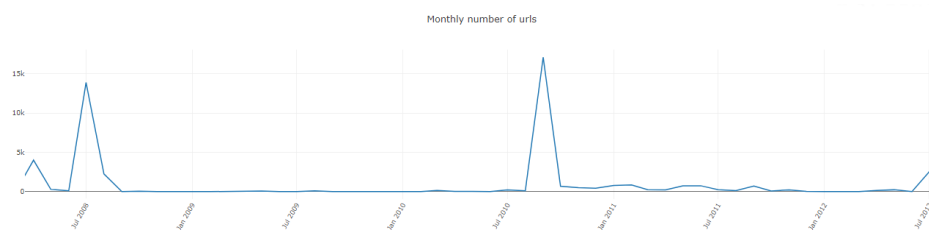


Figura 5.39: Métrica de conteo de enlaces, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.

La segunda métrica contabiliza las imágenes añadidas por los usuarios por meses.

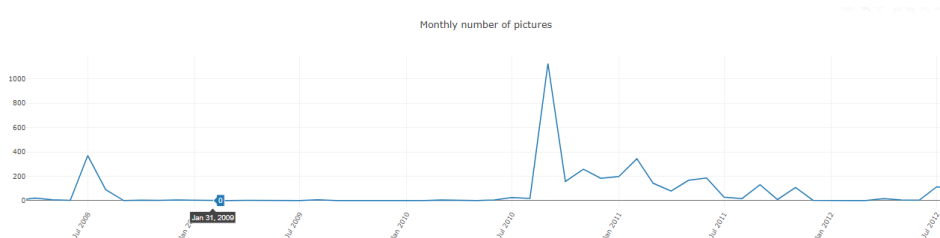


Figura 5.40: Métrica de conteo de imagenes, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.

La tercera métrica es de barras, con dos categorias enlace interno y enlace externo, nos parecía interesante ver la diferencia en cada wiki de enlaces a otras páginas o a la propia wiki.

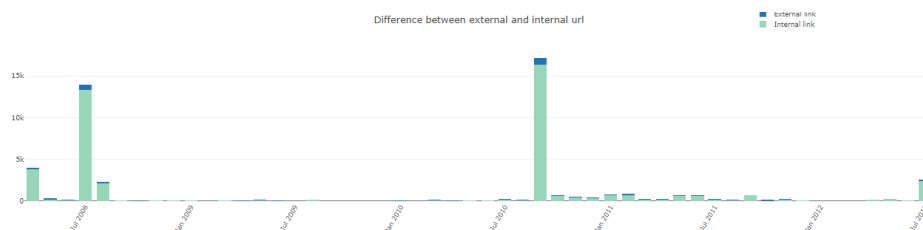


Figura 5.41: Métrica que muestra la división entre enlaces externos e internos, en la wiki de Cocktails, durante el periodo temporal de abril de 2008 a julio de 2012.

Capítulo 6

Análisis de wikis

6.1. Introducción

En este capítulo, ponemos en práctica algunas de las métricas de WikiChron Monowiki, con el objetivo de mostrar su utilidad, tanto para continuar estudios ya existentes sobre las wikis de Fandom, como para analizar nuevos fenómenos. Además, en algunos de estos estudios, se generan y demuestran hipótesis acerca de diversos factores de las comunidades colaborativas.

6.2. Tendencias estructurales opuestas en wikis de hobbies

En primer lugar, se proporciona, un estudio en el que se emplea WikiChron Monowiki para continuar la investigación comenzada en Serrano et al. (2018a). En dicha investigación, los autores se centran en el uso de WikiChron Classic, la versión previa a WikiChron Monowiki, que permite comparar varias wikis en el mismo gráfico, para estudiar fenómenos interesantes en algunas wikis de Fandom. Algunos de estos fenómenos, llegan a contradecir conclusiones generalizadas de estudios existentes centrados en comunidades más exitosas.

Para llevar a cabo esta investigación, emplean métricas que permiten estudiar la desigualdad en la participación de los contribuidores, como el coeficiente de Gini. Además, emplean métricas que permiten observar cómo es la relación entre el trabajo realizado por el 10 % de usuarios más activos, y el del 90 % de usuarios restante. Por otra parte, emplean ratios que permiten analizar la distribución de la participación dentro de la élite de la wiki, observando la relación entre el usuario más activo y el usuario en los percentiles 5, 10 y 20, y la relación entre los usuarios en los percentiles 5 y 10, y 10 y 20, respectivamente.

En el estudio expuesto a continuación, nos centramos, en concreto, en continuar su análisis para las wikis de Cocktails ¹ y Home ², durante sus primeros 120 meses. En ellas, de acuerdo con Serrano et al. (2018a), se observan tendencias de desigualdad de participación opuestas a lo largo del tiempo, de manera que, mientras ambas wikis crecen en cuanto a número de usuarios y número de ediciones, en una de ellas, la wiki de Cocktails, el coeficiente de Gini crece a lo largo del tiempo, mientras que en la otra, la wiki de Home, decrece, lo que implica que la participación en la comunidad de Home es menos inclusiva, es decir: hay un grupo de usuarios que hace la mayor parte del trabajo, y esta diferencia aumenta con el tiempo.

Al final del periodo estudiado, ambas wikis poseen un tamaño similar en cuanto a usuarios registrados y anónimos, de acuerdo con las observaciones de Serrano et al. (2018a). Además, la wiki de Cocktails posee un total de 1700 artículos, y la de Home, un total de 900 artículos.

En el estudio realizado a continuación, para completar esta investigación, se emplean distintas métricas de WikiChron Monowiki para desgranar y explicar mejor los fenómenos descritos en Serrano et al. (2018a), además de exponer cómo varía la estructura de la comunidad de usuarios a lo largo del tiempo. Para ello, se emplean las siguientes métricas del menú de Distribución de usuarios registrados activos:

- Active editors by experience (versión normal y en porcentaje), que permite determinar cómo de experimentados son los usuarios activos cada mes.
- Users by tenure (versión en porcentaje), que permite determinar la antigüedad de los usuarios activos en cada mes.
- Users by the date of the last edit (versión en porcentaje), que permite distinguir entre usuarios activos que vuelven a la wiki tras periodos de inactividad más o menos duraderos.

Además, se emplean las siguientes métricas del menú de Distribución de ediciones según el tipo de editor: Edits by editor experience (incluida versión en porcentaje) que permite estudiar cuál de las cuatro clases de usuarios activos por su experiencia editando, es más activo a lo largo del tiempo.

A continuación, se emplean las métricas de WikiChron Monowiki para explicar algunos de los fenómenos observados en Serrano et al. (2018a), además de para obtener una visión general de lo que sucede en ambas comunidades en sus primeros 120 meses.

¹https://cocktails.fandom.com/wiki/Cocktails_Wiki

²https://home.fandom.com/wiki/Home_Wiki

6.2.1. Cocktails

En la wiki de Cocktails, en torno al mes 87, la wiki experimenta un crecimiento en cuanto al número de ediciones, así como un aumento en la comunidad de editores que la conforma, de acuerdo con las observaciones de WikiChron Classic en Serrano et al. (2018a).

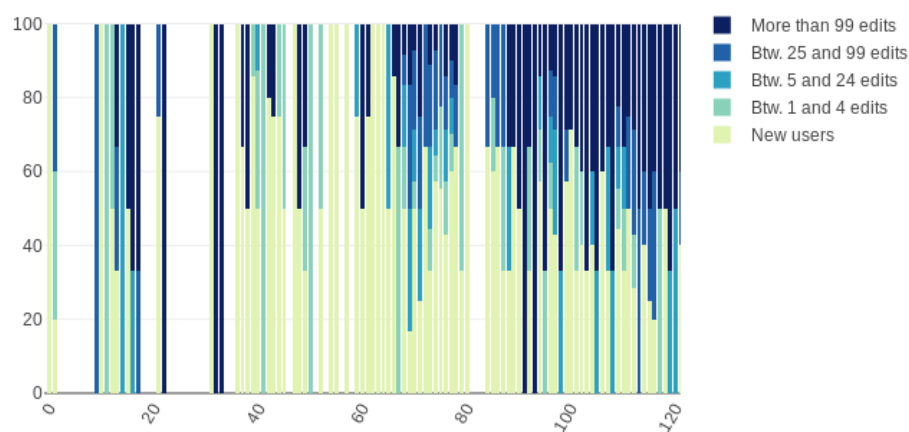


Figura 6.1: Métrica de Active registered users by experience (in percentage), en la wiki de Cocktails, durante sus primeros 120 meses.

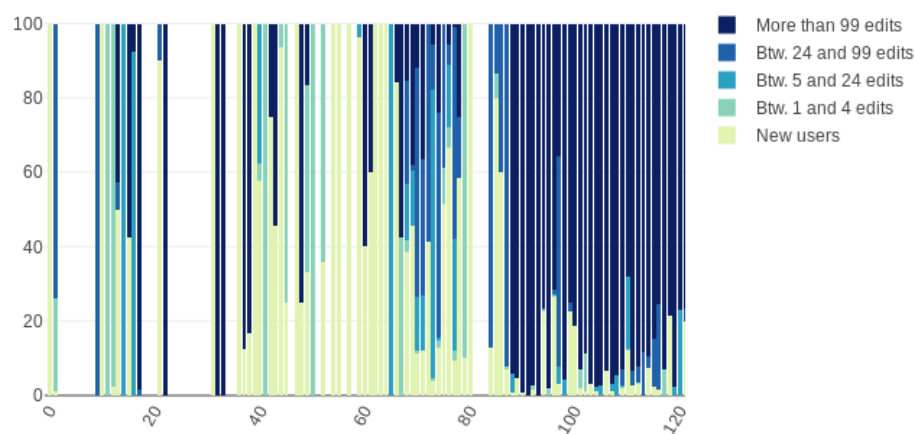


Figura 6.2: Métrica de Edits by editing experience (in percentage), en la wiki de Cocktails, durante sus primeros 120 meses.

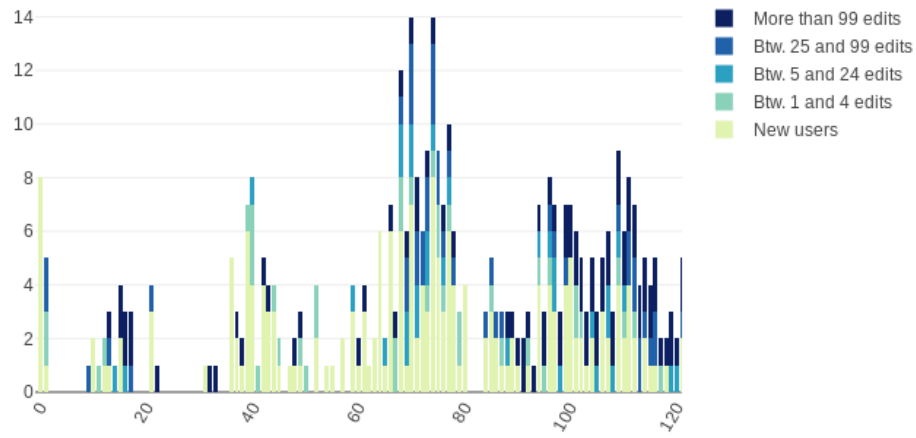


Figura 6.3: Métrica de Active registered users by experience, en la wiki de Cocktails, durante sus primeros 120 meses.

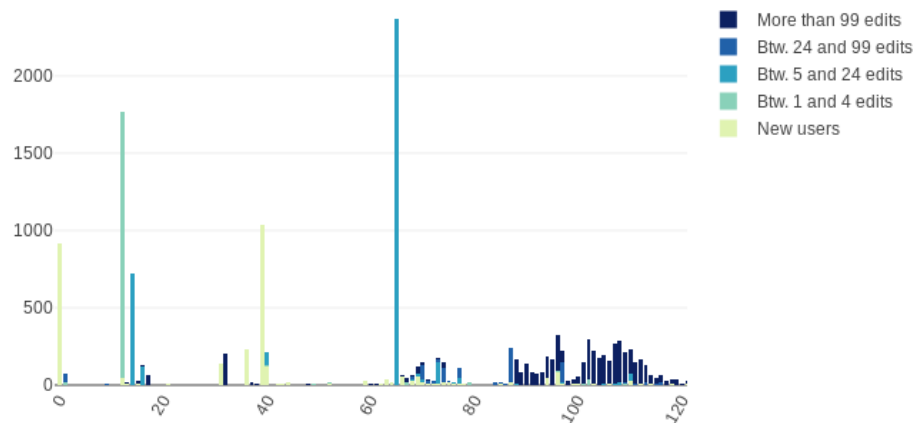


Figura 6.4: Métrica de Edits by editing experience, en la wiki de Cocktails, durante sus primeros 120 meses.

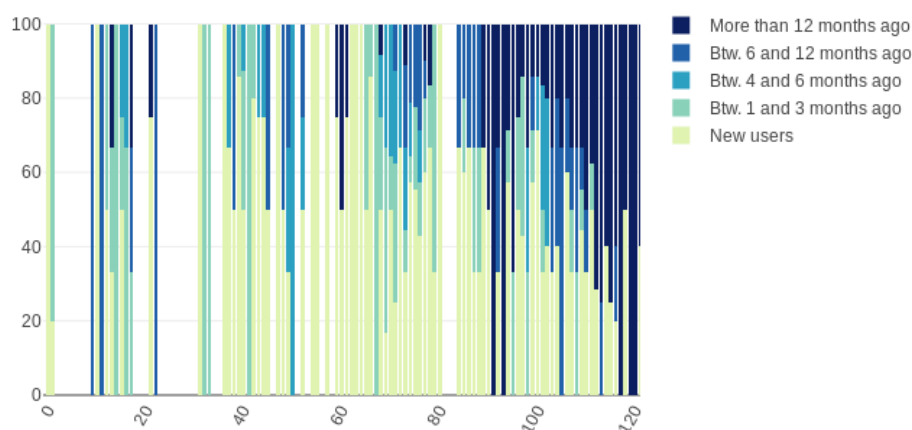


Figura 6.5: Métrica de Active registered users by tenure (in percentage), en la wiki de Cocktails, durante sus primeros 120 meses.

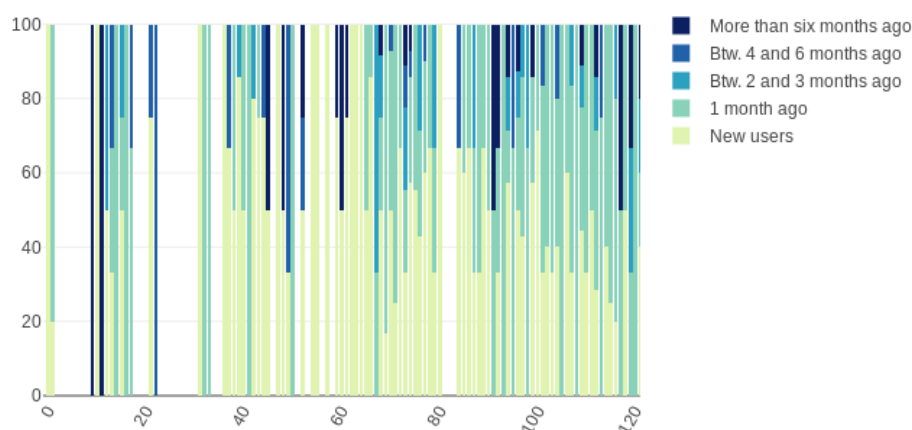


Figura 6.6: Métrica de Active registered users by the date of the last edit (in percentage), en la wiki de Cocktails, durante sus primeros 120 meses.

En torno a este mismo mes, hay un cambio en la soberanía de la wiki, que se observa claramente en la métrica de Edits by editor experience (versión en porcentaje), en la figura 6.2. Es decir, el grupo de usuarios más experimentados (aquellos que han realizado más de 99 ediciones), pasan a ser el grupo más prolífico, dejando en un segundo plano al grupo de usuarios nuevos, que de acuerdo con dicha métrica, hasta ese momento, eran los más colaboradores.

Teniendo en cuenta que esta métrica muestra qué porcentaje de ediciones, con respecto al total de ediciones en cada mes, lleva a cabo cada clase de usuarios, conviene observar la métrica de Active editors by experience y de Edits by editor experience en su versión normal, con el objetivo de discernir si la proporción de actividad de los usuarios con más de 99 ediciones hasta

el momento, aumenta al disminuir la cantidad de usuarios nuevos en la wiki, o al disminuir su actividad.

Así, tal y como se observa en la métrica de Active editors by experience (en ambas versiones), en las figuras 6.3 y 6.1, la cantidad de usuarios nuevos activos en cada mes, no disminuye con respecto a momentos anteriores al mes 87, mientras que la participación de usuarios experimentados activos sí se hace más intensa de lo que lo era en meses anteriores.

Por otra parte, en la métrica de Edits by editor experience (en su versión sin porcentaje), mostrada en la figura 6.4, se observa que la actividad del grupo de usuarios nuevos disminuye ligeramente con respecto a momentos anteriores, aunque no lo suficiente como para justificar por sí misma el cambio de soberanía observado con la métrica de Edits by editor experience en su versión en porcentaje, mostrada en la figura 6.2. Además, se observa que la actividad de los usuarios experimentados aumenta. Así, este cambio de soberanía se puede considerar una consecuencia directa de que los usuarios experimentados se hacen más prolíficos en la wiki, ya que colaboran en una mayor cantidad de ediciones y en una mayor constancia. Además, este fenómeno también se ve impulsado por una ligera disminución de la cantidad de ediciones producidas por el grupo de usuarios nuevos.

Las métricas de WikiChron Monowiki proporcionan una explicación más profunda y en términos de estructura de la comunidad para las observaciones con WikiChron Classic en Cocktails por Serrano et al. (2018a): el aumento en la actividad de la wiki (en términos de número de usuarios y de ediciones), en torno al mes 87, observado en dicho estudio, se explica por una mayor actividad de los usuarios más experimentados de la wiki.

Esta actividad está causada tanto por una presencia más constante de este grupo de usuarios con más de 99 ediciones, y por un aumento en su colaboración, además de que el grupo de usuarios activos nuevos, que era el más numeroso, deja de serlo a favor de los usuarios con más de 99 ediciones de experiencia.

Por otra parte, surge la pregunta de si los usuarios muy experimentados que surgen en la wiki a partir del mes 87 son usuarios que rápidamente alcanzan el estatus de experimentado, o si se trata de usuarios más antiguos que poco a poco van acumulando ediciones hasta llegar a tener el estatus de más de 99 ediciones.

Con la métrica de Users by tenure, en la figura 6.5, se observa que, a partir del mes 87, la mayor parte de los usuarios activos son usuarios que llegaron a la wiki hace más de 12 meses: antes del mes 74. Por tanto, a partir del mes 87, la comunidad está formada por un grupo de usuarios muy experimentados en cuanto al número de ediciones en el que contribuyen, y que han llegado a la wiki en un momento anterior al mes 74.

Además, estos usuarios son muy constantes, lo que se observa en la métrica de Users by the date of the last edit, en la figura 6.6, en la que se aprecia que a partir del mes 67, comienza un período en el que los usuarios activos

son sobre todo, usuarios cuya última edición en la wiki fue el mes anterior.

Por tanto, los usuarios activos tan experimentados en cuanto a experiencia editando que son los más contribuyentes a partir del mes 87, son usuarios que llegan antes del mes 74 y que editan con mucha constancia desde su llegada a la wiki.

En definitiva, con estas últimas observaciones, hemos podido concluir que la wiki de Cocktails comienza a retener y a comprometer a más usuarios a partir del mes 67, momento en el que los usuarios activos son, sobretodo, usuarios con más de 12 meses de antigüedad, y que contribuyen en la wiki todos los meses. Estos usuarios, a partir del mes 87, alcanzan el estatus de muy experimentados, llegando a formar parte del grupo de veteranos de la wiki.

6.2.2. Home

La wiki de Home, no obstante, crece de forma más lenta que la wiki de Cocktails, de acuerdo con las observaciones en Serrano et al. (2018a).

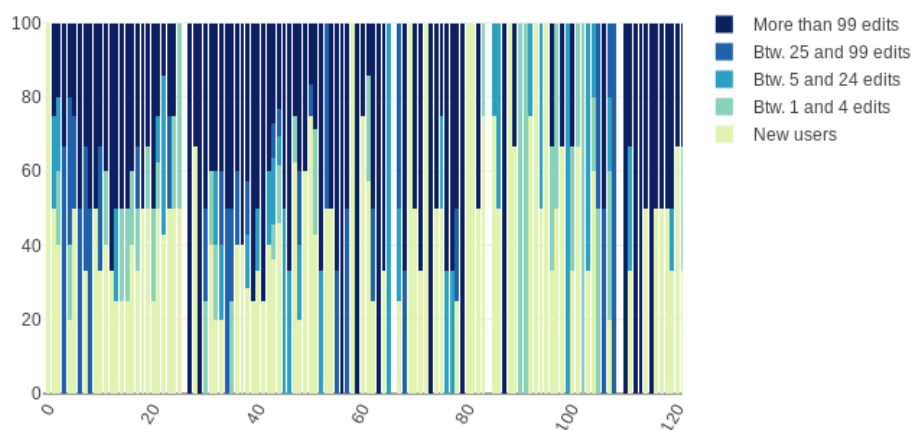


Figura 6.7: Métrica de Active registered users by editing experience (in percentage), en la wiki de Home, durante sus primeros 120 meses.

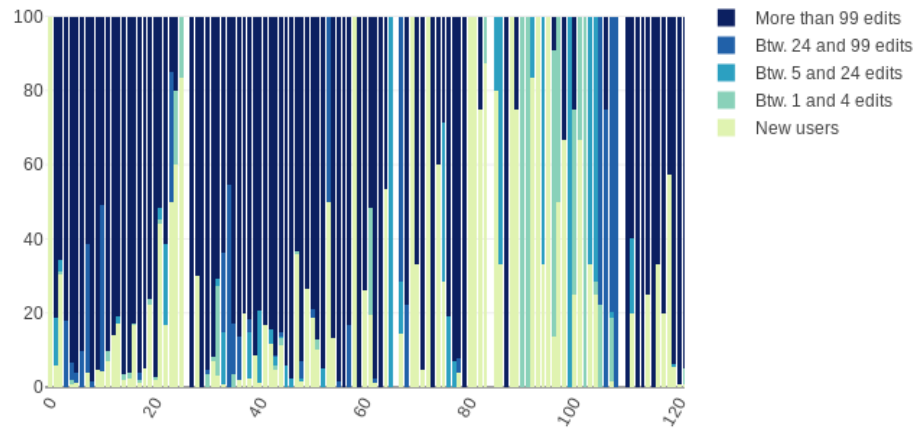


Figura 6.8: Métrica de Edits by editing experience (in percentage), en la wiki de Home, durante sus primeros 120 meses.

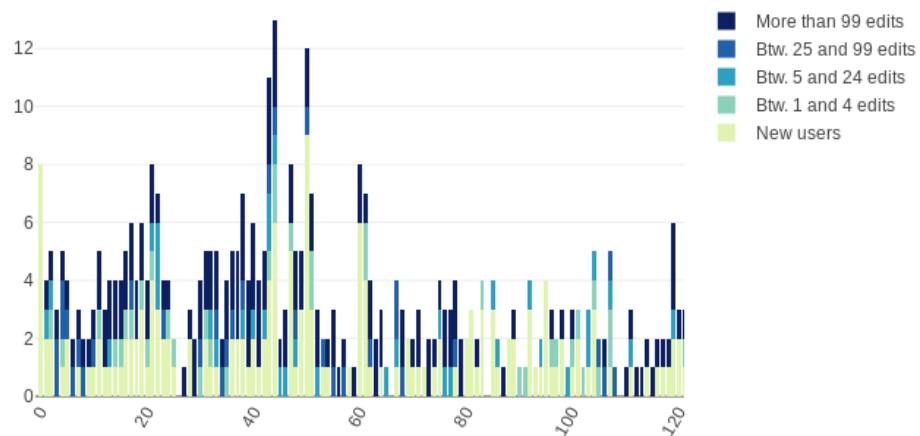


Figura 6.9: Métrica de Active registered users by experience, en la wiki de Home, durante sus primeros 120 meses.

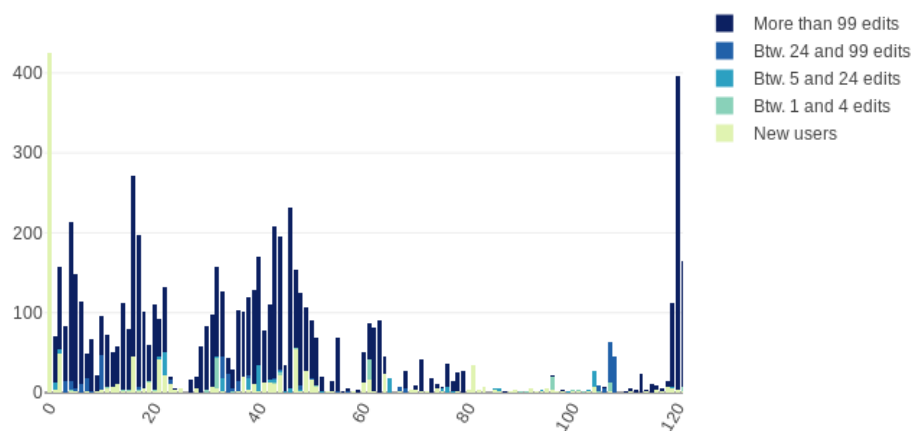


Figura 6.10: Métrica de Edits by editing experience, en la wiki de Home, durante sus primeros 120 meses.

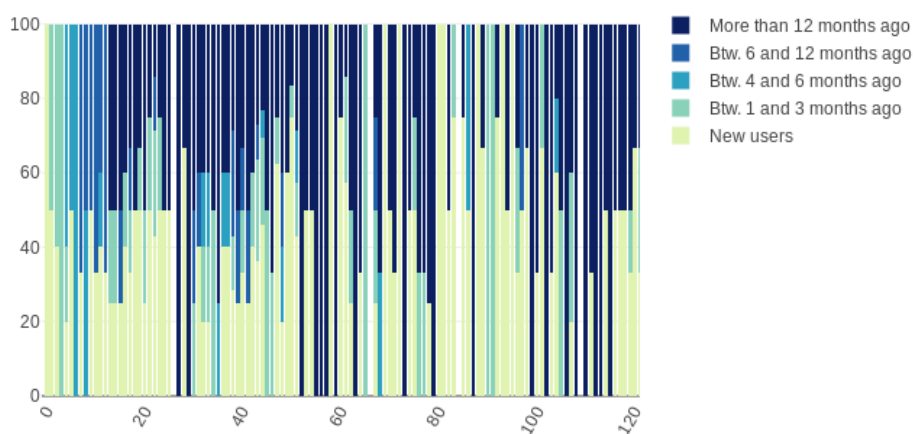


Figura 6.11: Métrica de Active registered users by tenure (in percentage), en la wiki de Home, durante sus primeros 120 meses.

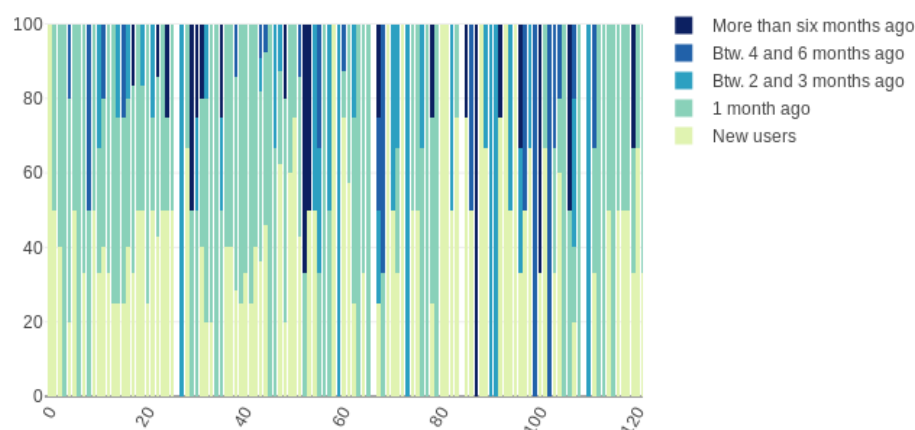


Figura 6.12: Métrica de Active registered users by the date of the last edit (in percentage), en la wiki de Home, durante sus primeros 120 meses.

Con la métrica de Edits by editor experience en su versión en porcentaje, mostrada en la figura 6.8, se observa que, a diferencia de Cocktails, se trata de una wiki en la que el grupo de editores más prolífico está formado por usuarios experimentados (usuarios que han llevado a cabo más de 99 ediciones hasta el momento) desde el principio de la wiki.

No obstante, este patrón varía en torno al mes 80, momento en el que ocurre un giro: el grupo más prolífico pasa a ser el grupo de usuarios nuevos.

Dado que el valor de esta métrica es un porcentaje con respecto al total de ediciones en cada mes, conviene observar la métrica de Active editors by experience y la de Edits by editor experience en relativo, con el objetivo de clarificar si la proliferación de la clase de usuarios nuevos como clase más activa es consecuencia de un abandono por parte de los experimentados, o si, por el contrario, es consecuencia de una mayor cantidad de usuarios nuevos, o de la misma cantidad de usuarios nuevos editando más.

La figura 6.9 muestra los resultados de la métrica Active editors by experience, en la que se observa que, a partir del mes 80, hay un ligero aumento en la cantidad de usuarios nuevos que llegan cada mes a la wiki. Además, entre el mes 80 y el mes 110, se observa que hay una menor presencia del grupo de usuarios que han realizado más de 99 ediciones hasta el momento. Algún miembro de este grupo reaparece ocasionalmente, pero su constancia, en este periodo temporal, es mucho menor.

Por otra parte, la métrica de Edits by editor experience en su versión sin porcentaje, en la figura 6.10, muestra que, a partir del mes 80, el grupo de usuarios nuevos, ahora algo más amplio, no contribuye en una mayor cantidad de ediciones en comparación con su actividad anterior. Así, este cambio de soberanía, por el que, a partir del mes 80, la clase de usuarios nuevos pasa a ser la más prolífica, se explica por un abandono del grupo de usuarios experimentados, cuya actividad se vuelve menor y menos constante.

Este patrón se mantiene hasta el mes 110, momento en el que, como se observa en la métrica de Active editors by experience, en la figura 6.9, los usuarios experimentados vuelven a la wiki.

A partir de esta última observación surge una nueva pregunta: ¿son estos usuarios veteranos nuevos, o se trata de los antiguos usuarios veteranos que abandonaron la wiki en torno al mes 80 y que vuelven en torno al mes 110?

En la métrica de users by tenure, en la figura 6.11, se ve que a partir del mes 100 (incluido el mismo), los usuarios más prolíficos son usuarios cuya primera edición en la wiki fue hace más de 12 meses. Es decir, se trata de usuarios cuya primera aparición en la wiki data de antes de antes del mes 85. Esto puede llevar a pensar que los usuarios tan experimentados en cuanto al número de ediciones que aparecen en el mes 110 (figura 6.7), son los mismos que abandonaron la wiki en torno al mes 80.

Sin embargo, en la métrica de Users by the date of the last edit, en la figura 6.12, se observa que en el período temporal que parte del mes 100, después de los usuarios nuevos, los usuarios más prolíficos son aquellos cuya última edición en la wiki ha sido hace 1 mes, seguidos de aquellos usuarios cuya última edición fue hace 4 y 6 meses y hace 2 y 3 meses. Por tanto, los usuarios cuya primera edición fue hace más de 12 meses que se muestra en la gráfica de Users by tenure a partir del mes 100, no son los veteranos que llevaban a cabo la mayor parte de las contribuciones de la wiki antes del mes 80, sino usuarios que aparecen a partir del mes 80 y en adelante.

6.2.3. Conclusiones

A continuación, se emplean las observaciones detalladas en las secciones 6.2.1 y 6.2.2 para proporcionar una visión más profunda que permite explicar mejor los fenómenos observados en Serrano et al. (2018a).

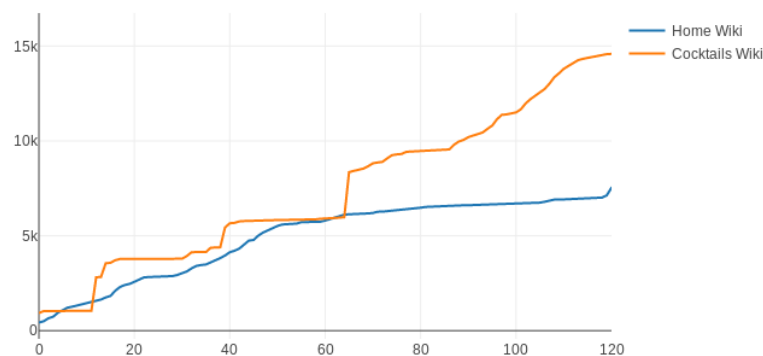


Figura 6.13: Métrica de Total edits in pages de WikiChron Classic, en las wikis de Home (naranja) y Cocktails (azul), durante sus primeros 120 meses.

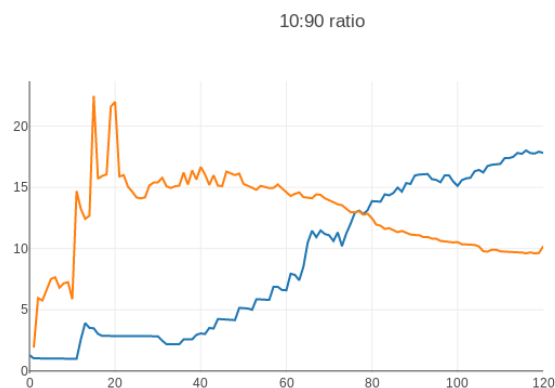


Figura 6.14: Métrica del Ratio 10:90, en las wikis de Home (naranja) y Cocktails (azul), durante sus primeros 120 meses.

En la **wiki de Cocktails**, hay un claro cambio de soberanía en torno al mes 87, momento en el que se abren paso los usuarios experimentados como grupo más prolífico. Este cambio va de la mano de una participación más activa de este grupo de usuarios, junto con un aumento de su presencia en la wiki. Este cambio se ve reflejado, además, en el crecimiento de las ediciones en la wiki a partir de dicho mes, mostrado al observar la métrica de Total edits in pages de WikiChron Classic, en la figura 6.13.

No obstante, al observar la métrica de WikiChron Classic del ratio 10:90, mostrada en la figura 6.14, este cambio de soberanía no tiene ningún impacto especial en el ratio 10:90, que se mantiene creciente durante gran parte del tiempo de vida de la wiki. Esto sugiere que durante el cambio de soberanía, los usuarios experimentados sustituyen rápidamente a aquellos usuarios situados en el 10 % de usuarios más activos. Este hecho va de la mano de la métrica de edits in pages de WikiChron Classic, en la figura 6.13, que a partir del mes 87, crece notablemente con respecto a su valor en meses anteriores.

Este fenómeno se considera una evidencia más de la hipótesis de élite inclusiva, expuesta en el artículo Serrano et al. (2018a), dado que, a pesar del cambio de patrón observado entre usuarios experimentados y usuarios nuevos, nunca hay un bajón en el ratio 10:90.

Por otra parte, en la **wiki de Home**, se observa lo contrario. Desde el inicio de la wiki, se observa que el grupo de usuarios más prolífico es el de los experimentados, es decir, aquellos que han llevado a cabo más de 99 ediciones hasta el momento. No obstante, en torno al mes 80, este patrón cambia: los usuarios experimentados parecen abandonar la wiki.

A partir de ese momento, el grupo de usuarios más prolífico es, por primera vez en esta wiki, el grupo formado por los usuarios nuevos que llegan

a la wiki cada mes. Este abandono por parte de los experimentados, se hace notar tanto en el número de ediciones, observado en la métrica de WikiChron Classic Total edits in pages, en la figura 6.13, como en el ratio 10:90 de WikiChron Classic, mostrado en la figura 6.14.

Este último ratio, comienza a decrecer de forma más rápida a partir del mes 80, momento que coincide con el abandono por parte de los experimentados, para volver a crecer en torno al mes 118, coincidiendo con el momento en el que vuelve a crecer el número de ediciones de la wiki, y el momento en el que vuelve a haber usuarios experimentados en la wiki que comienzan a ser más prolíficos.

Esta observación lleva a pensar que Home es una wiki en la que la élite está estrictamente formada por un grupo de usuarios experimentados, que, al desaparecer, provoca una caída del ratio 10:90. Esto implica que, a partir del mes 80 y hasta en torno al momento en el que vuelven los experimentados, en el mes 110, el grupo más prolífico, formado por usuarios nuevos, no logra formar parte del 10 % de los usuarios más activos, sugiriendo que este grupo de usuarios nuevos no contribuye lo suficiente como para lograr llegar a la élite, formada por los usuarios experimentados.

Además de este decrecimiento del ratio 10:90, las observaciones en las métricas de Edits by editor experience, en la figura 6.10, y Active editors by experience, en la figura 6.9, sugieren que ni el número de usuarios nuevos ni sus contribuciones aumentan sustancialmente en comparación con momentos anteriores de la wiki, por lo que la actividad del grupo de usuarios nuevos no varía, sino que su porcentaje de actividad con respecto al resto se hace más grande al desaparecer los experimentados.

Así, la actividad que llevan a cabo estos usuarios en ausencia de los experimentados, no es suficiente como para que estos usuarios nuevos pasen a formar parte de la élite de Home, y evitar la bajada del ratio 10:90 que se produce.

En definitiva, este estudio aporta evidencias estructurales a las conclusiones hechas por Serrano et al. (2018a), además de demostrar que un cambio de soberanía en el grupo de usuarios más activos de la wiki, no implica un cambio en la élite en la misma.

6.3. Análisis del compromiso de los usuarios en distintas wikis

Tenemos cinco métricas dentro de nuestro wikichron monowiki enfocadas en analizar el compromiso de los usuarios. Para ello vamos a coger dos wikis que son interesantes para ver este tipo de análisis.

6.3.1. Laguna negra (El Internado)

Lo que buscamos analizando esta wiki, es ver la relacion que hay entre el compromiso de los usuarios y la actividad de la serie de televisión. Lo primero de todo es explicar el contexto de la wiki.

Laguna negra es una wiki creada para hablar sobre la serie de televisión de El Internado. Esta serie se emitió por primera vez en mayo de 2007 y finalizó en octubre de 2010 con un total de siete temporadas, emitiendo, cada año, una temporada entera y parte de alguna otra. La wiki se creó en abril del 2009, coincidiendo con la noticia del estreno de la quinta temporada.

Esta wiki fue creada por un solo individuo, ya que solo hay un usuario que editó en el mes de su comienzo. Este individuo, intentó levantar un poco la wiki editando dos meses seguidos, pero no tuvo mucho éxito, ya que solo aparecieron tres usuarios nuevos que editaron este segundo mes y no volvieron a editar más en los meses posteriores. Esto lo podemos ver en la imagen 6.15.

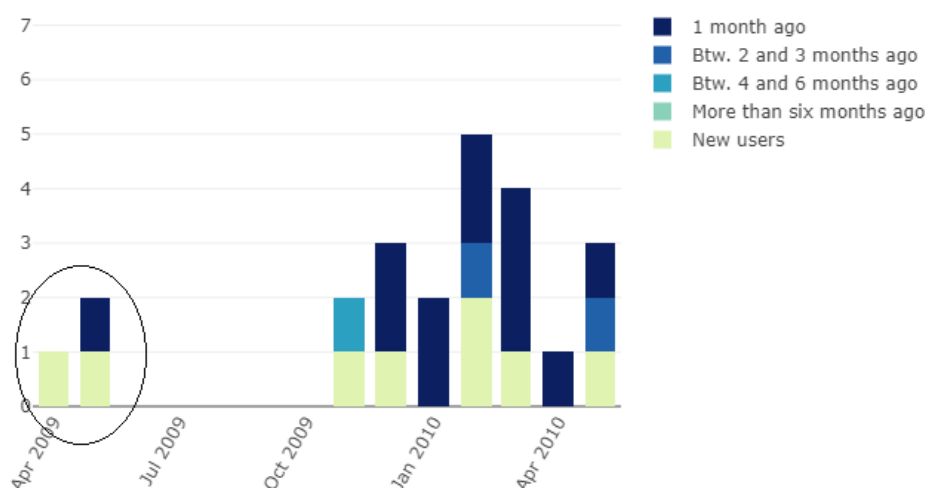


Figura 6.15: Métrica de usuarios activos registrados según su última edición, en la wiki de Laguna negra, desde abril de 2009 hasta mayo de 2010

Como la wiki, en un principio, no tuvo mucho éxito, el usuario que la creó se debió de cansar y dejó de realizar contribuciones a partir del tercer mes.

A pesar de este fatídico comienzo, podemos observar en la imagen 6.16, que en el mes de octubre de 2009 llega un nuevo usuario que realiza alguna edición y ya es en el mes siguiente cuando la wiki empieza a crecer en número de usuarios. A partir de noviembre empezamos a observar, como algunos usuarios que hicieron su primera edición hace ya algunos meses vuelven a

editar, es decir, se empiezan a enganchar a la wiki.

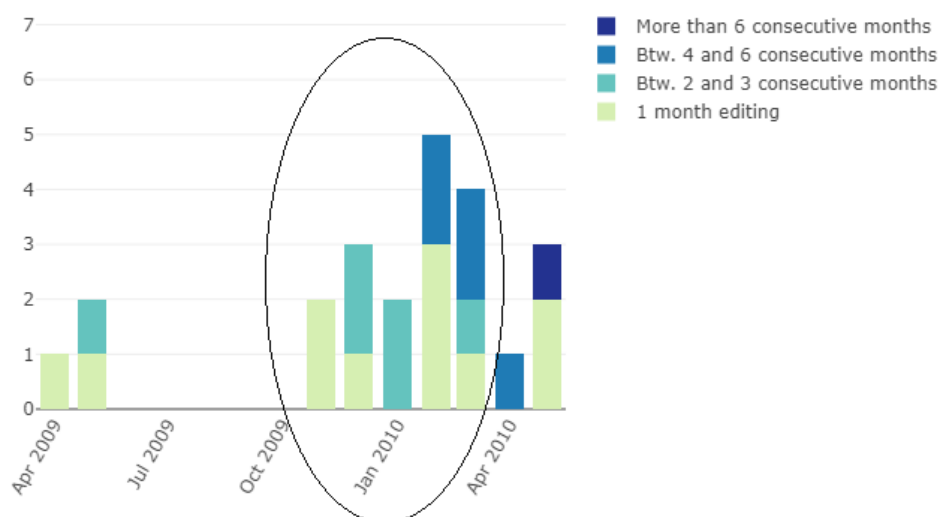


Figura 6.16: Métrica de usuarios activos registrados según su racha actual, en la wiki de Laguna negra, desde abril de 2009 hasta abril de 2010

En la figura 6.17 se puede apreciar como algunos contribuidores ya empiezan a comprometerse con la wiki y empiezan a realizar ediciones de forma frecuente. Cabe destacar que el que ahora lleva la wiki, es decir, el que edita todos los meses, no es el mismo usuario que la creó. Es un usuario que se integró en noviembre del 2009.

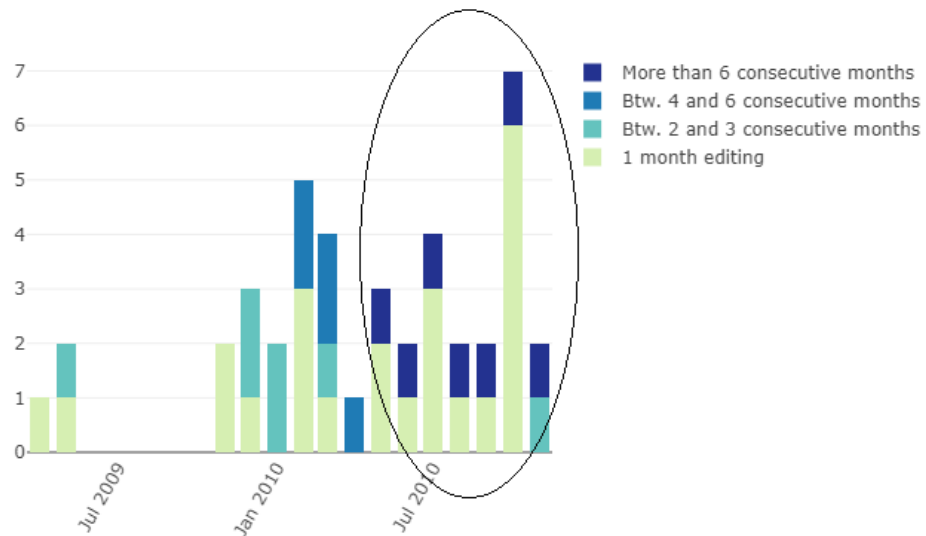


Figura 6.17: Métrica de usuarios activos registrados según su racha actual, en la wiki de Laguna negra, desde mayo de 2009 hasta noviembre de 2010

En diciembre de 2009 y febrero, julio y octubre de 2010 se aprecia, en la figura 6.18, como hay usuarios nuevos que les interesa demasiado la wiki y editan al menos dos veces en la misma semana de su registro.

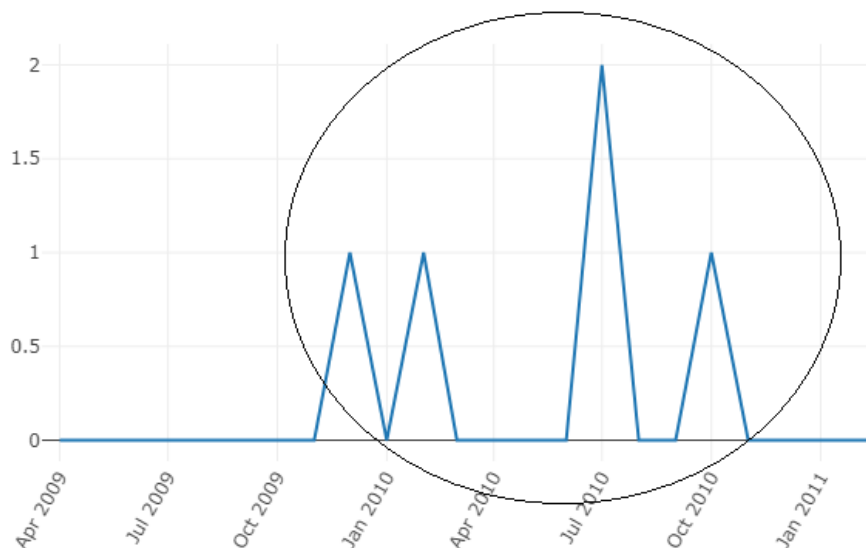


Figura 6.18: Métrica de nuevo usuario reincidente, en la wiki de Laguna negra, desde abril de 2009 hasta enero de 2011

Pero, en julio, los dos individuos que empiezan con muchas ganas, se

cansan rápido, ya que al mes siguiente ya no editan ninguno de los dos, como podemos ver en la figura 6.17, ya que no hay ningún usuario que haya editado dos meses consecutivos.

En octubre del 2010, hay una gran subida de usuarios nuevos, ya que coincide con el mes en el que se emite el último capítulo de la serie. Y además de usuarios nuevos, hay un reenganche de algunos usuarios que ya habían editado en algún momento, pero no habían vuelto a editar en, al menos, dos o tres meses previos, como se puede ver en la figura 6.19.

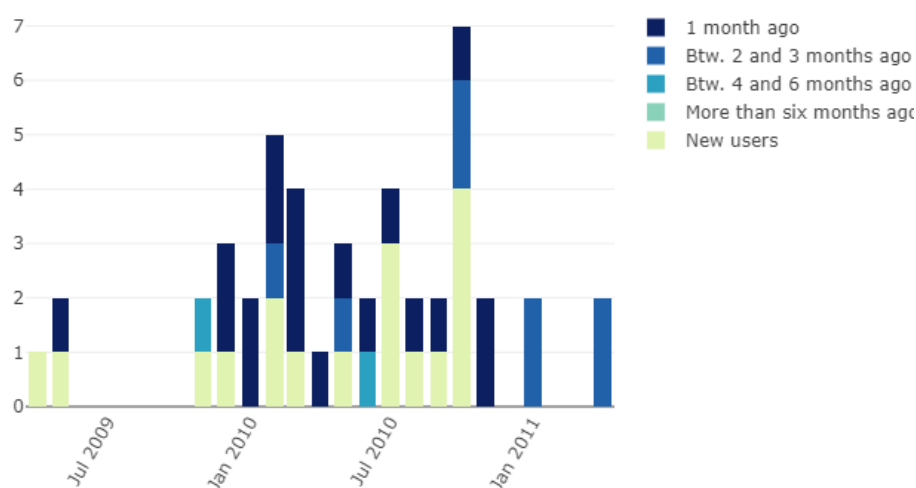


Figura 6.19: Métrica de usuario activos registrados según su última edición, en la wiki de Laguna negra, desde abril de 2009 hasta abril de 2011

En la imagen 6.20 vemos como al mes siguiente desciende de forma desmesurada el número de usuarios nuevo, aunque los existentes siguen siendo fieles. A partir del mes de diciembre, la wiki va perdiendo contribuidores, incluso el editor más fiel es en este mes cuando deja de editar.

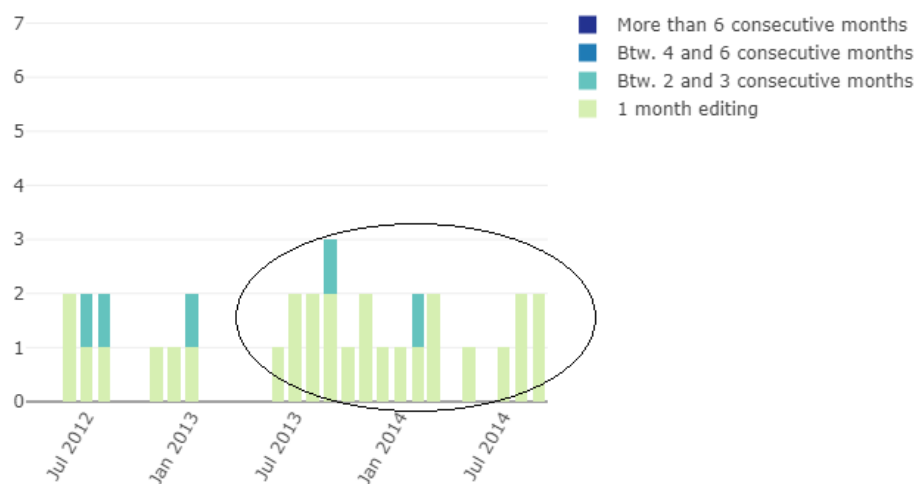


Figura 6.21: Métrica de usuarios activos registrados según su racha actual, en la wiki de Laguna negra, desde mayo de 2012 hasta septiembre de 2014

Por último, podemos destacar en la imagen 6.22 una nueva subida de usuarios nuevos y la aparición de editores que realizan su última edición hace más de seis meses sobre mediados o finales de 2017, que justamente coincide con el estreno del reencuentro de El Internado por el décimo aniversario del comienzo de la serie. Pero estos usuarios no están muy comprometidos con la wiki ya que como podemos ver en la imagen 6.23, como su racha actual, solo dura ese mes.

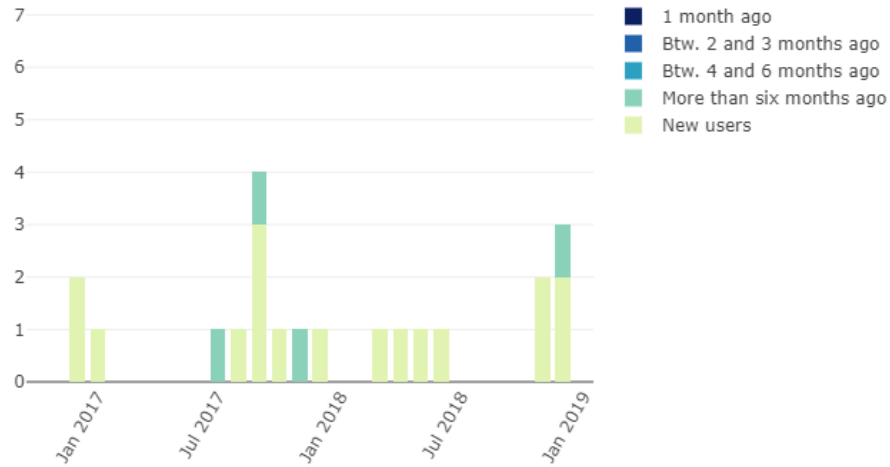


Figura 6.22: Métrica de usuarios activos registrados según su última edición, en la wiki de Laguna negra, desde octubre de 2016 hasta enero de 2019

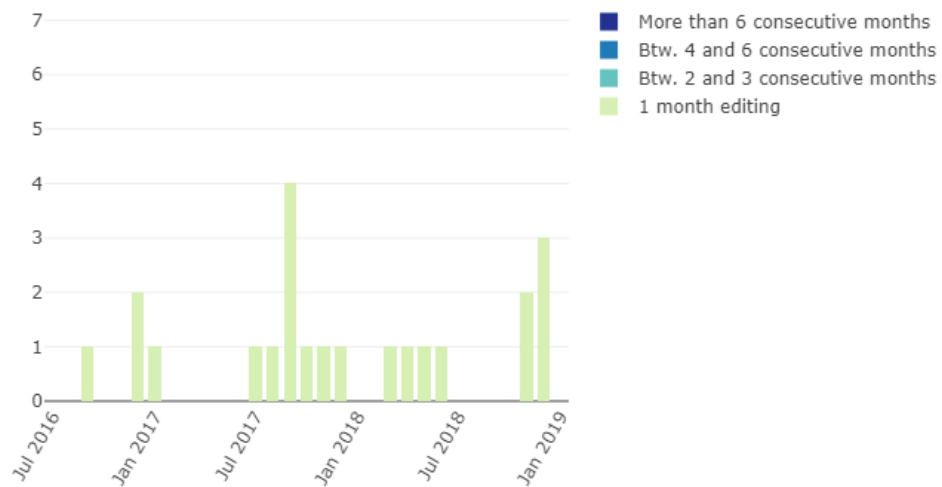


Figura 6.23: Métrica de usuarios activos registrados según su racha actual, en la wiki de Laguna negra, desde julio de 2016 hasta enero de 2019

Como conclusión, podemos decir, que generalmente, cuando hay un suceso en esta serie de televisión, este repercute sobre la wiki en cuestión. En este caso concreto, el compromiso de los usuarios se concentra en momentos en los que la serie está todavía en emisión, por el contrario, cuando se hace algún estreno en otro país o simplemente se vuelve a reproducir la serie,

el compromiso ya no es tan elevado, solo se aprecian usuarios de paso, que hacen ediciones en un momento determinado pero luego no continúan.

Una vez explicado el compromiso de los usuarios para esta wiki, vamos a ver como se distribuyen las ediciones entre ellos para determinar si son los usuarios más experimentados los que realizan la mayor parte de las ediciones, o, por el contrario, los más novatos.

En general, como podemos ver en las imágenes 6.24, 6.25 y 6.26, en el primer año y medio de la vida de la wiki, se observa como la mayor parte de las ediciones son realizadas por usuarios bastante comprometidos generalmente. Esto tiene sentido, ya que es normal que estos usuarios más comprometidos, se preocupen más por la wiki. En cambio, en octubre de 2010, pasa todo lo contrario, es decir, la mayor parte de las ediciones son realizadas por usuarios nuevos o muy poco comprometidos, esto se debe a que, como hemos comentado antes, este es el mes en el que la serie llega a su fin en España y hay un despunte de usuarios nuevos que tienen mucha motivación con la serie.

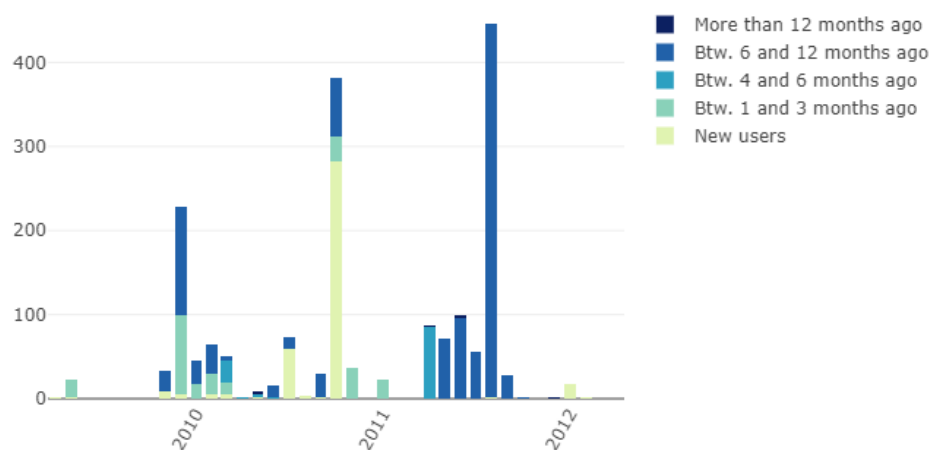


Figura 6.24: Métrica de dist. ediciones por usuarios activos registrados según su antigüedad, en la wiki de Laguna negra, desde abril de 2009 hasta abril de 2012

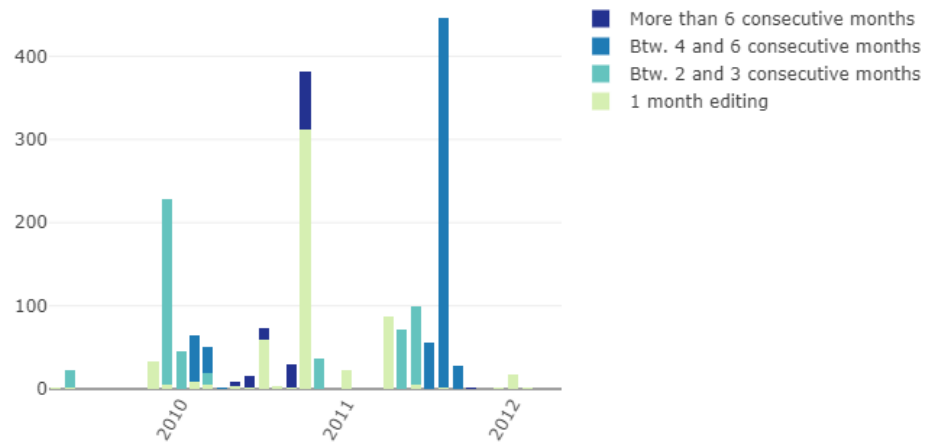


Figura 6.25: Métrica de dist. ediciones por usuarios activos registrados según su racha actual, en la wiki de Laguna negra, desde abril de 2009 hasta abril de 2012

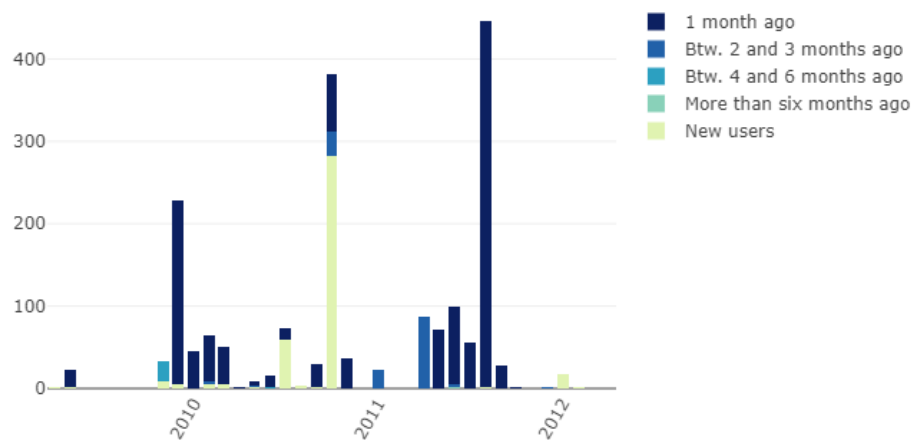


Figura 6.26: Métrica de dist. ediciones por usuarios activos registrados según su última edición, en la wiki de Laguna negra, desde abril de 2009 hasta abril de 2012

Una cosa interesante que también se observan en las figuras 6.24, 6.25 y 6.26 es que en agosto de 2011, casi todas las ediciones realizadas en ese mes están hechas por usuarios muy comprometidos, pero si vemos, por ejemplo, la imagen 6.20 observamos como solo es un usuario el que hace ese porrón de ediciones. Esto nos puede indicar o que ha habido un acto vandálico y que este usuario se haya dado cuenta y lo esté arreglando o que es un currante

que está cambiando la wiki de arriba a abajo.

Por lo tanto, podemos concluir diciendo que normalmente, los usuarios que tienen más compromiso dentro de la wiki son también aquellos que más trabajan, es decir, que más ediciones realizan. Esto puede dejar de ser así cuando ocurre algún evento en la vida real que repercuta en la wiki y en ese caso, los usuarios más nuevos también se motivan.

6.3.2. WikiAnswer Pets

Es una wiki de consulta, es decir, algunos usuarios formulan preguntas relacionadas con las mascotas, y otros responden a esas preguntas. Como vemos en la figura 6.27 fue creada en junio de 2007 por dos usuarios.

Vemos como en el primer año y medio de la wiki, no hay apenas compromiso de los usuarios, ya que ni siquiera hay usuarios activos como podemos ver en la figura 6.27.

Es a partir de enero de 2009, cuando esta wiki comienza a crecer.

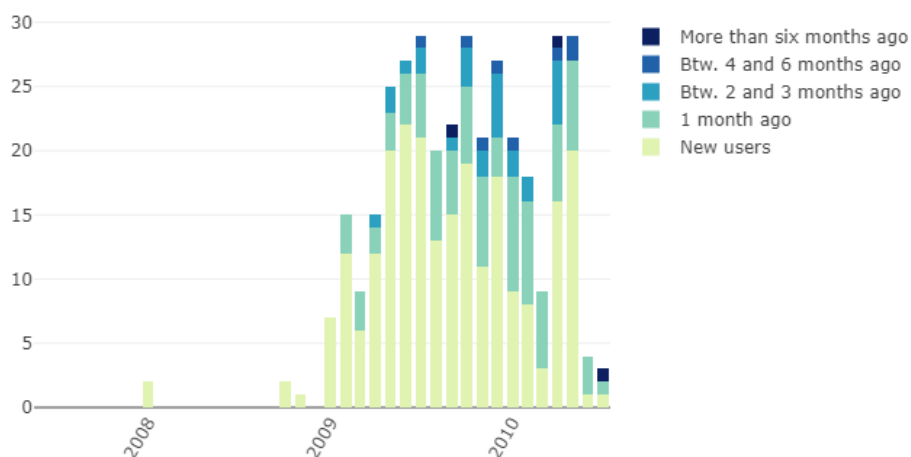


Figura 6.27: Métrica de usuarios activos registrados según su última edición, en la wiki de Wikianswer Pets, desde junio de 2007 hasta julio de 2010

Vemos, como, a raíz de este mes, la wiki va cogiendo importancia, y vemos incluso como hay algunos usuarios que han vuelto a editar pasando algunos meses. Los usuarios nuevos que aparecen en esta época empiezan con ganas, ya que hay muchos de ellos que editan varias veces dentro de la semana en la que se registran, aunque suele pasar que empiezan con muchas ganas y luego se olvidan como podemos ver en la figura 6.28 y 6.29.

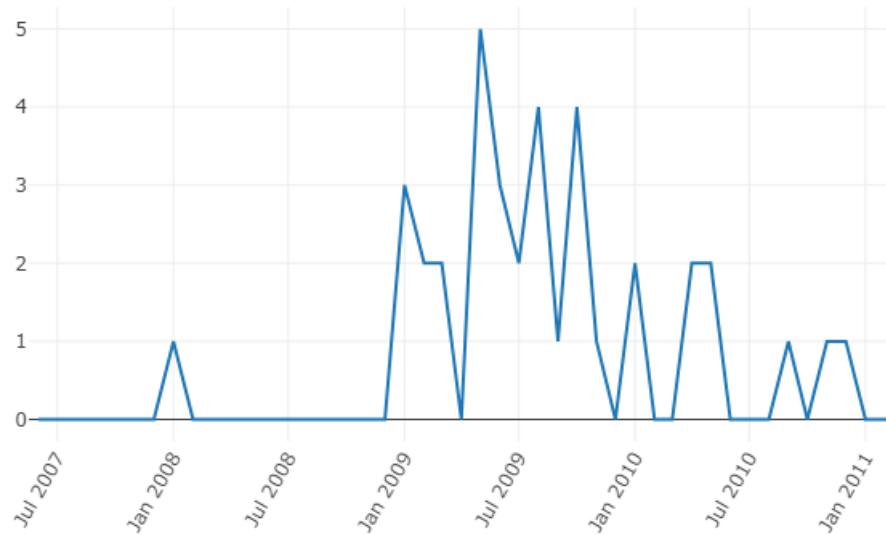


Figura 6.28: Métrica de nuevo usuario reincidente, en la wiki de Wikianswer Pets, desde julio de 2007 hasta enero de 2011

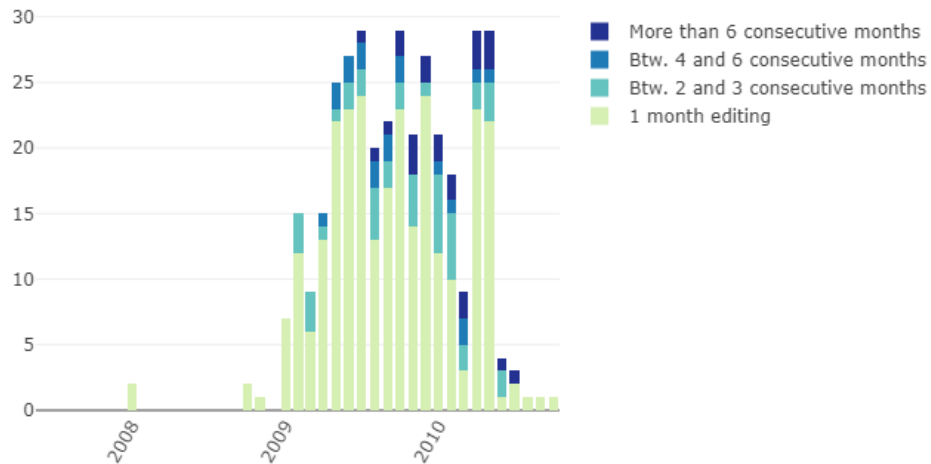


Figura 6.29: Métrica de usuarios activos registrados según su racha actual, en la wiki de Wikianswer Pets, desde julio de 2007 hasta octubre de 2010

Por ejemplo, en mayo, hay 5 editores que comienzan con ganas, pero vemos como al mes siguiente solo han editado dos usuarios entre dos y tres meses, así que como máximo, de esos cinco, solo puede haber dos que hayan editado después. Por otro lado, vemos como en este momento, hay bastantes usuarios que editan frecuentemente.

Como vemos en la imagen 6.30, a partir de junio, la wiki empieza a decrecer en número de usuarios y también decrece el compromiso. En abril de 2011 el número de usuarios nuevos aumenta, pero, el compromiso sigue siendo bajo, como vemos en la figura 6.31. En meses posteriores, la wiki, recibe usuarios nuevos de forma más o menos constante. Y es ya, entre finales de 2017 y 2018, cuando el número de estos usuarios comienza a bajar y el compromiso es demasiado escaso.

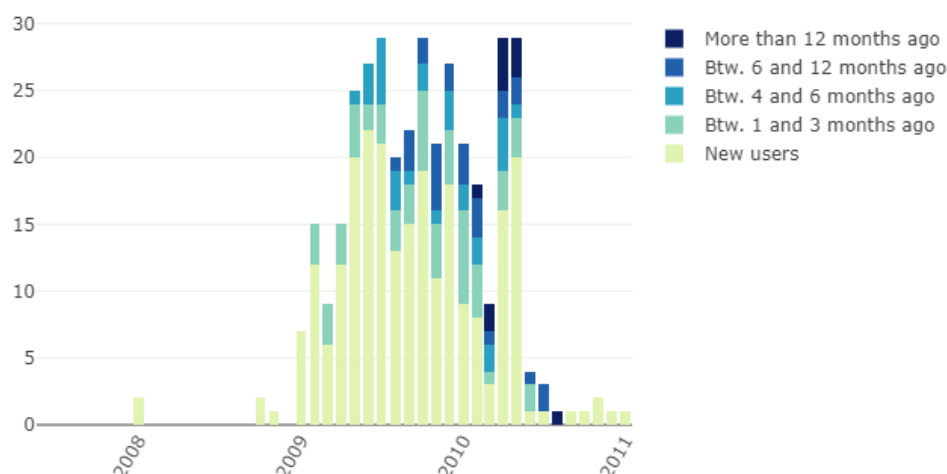


Figura 6.30: Métrica de usuarios activos registrados según su antigüedad, en la wiki de Wikianswer Pets, desde julio de 2007 hasta enero de 2011

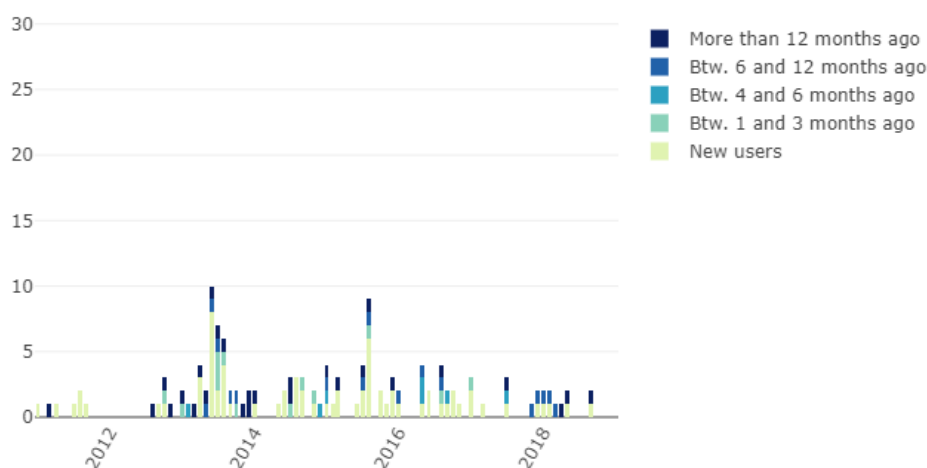


Figura 6.31: Métrica de usuarios activos registrados según su antigüedad, en la wiki de Wikianswer Pets, desde enero de 2011 hasta enero de 2019

En general, los usuarios de esta wiki están muy comprometidos entre 2009 y 2010 y esto coincide con la presencia de un elevado número de usuarios nuevos, por lo que estos son bastante fieles. Por el contrario, posteriormente, el compromiso decrece, es decir, los usuarios que estaban muy comprometidos ya se han cansado y los usuarios nuevos se cansan rápido también. Podemos ver también una relación entre número de usuarios nuevos y el compromiso de ellos, ya que cuanto mayor número de usuarios, mayor probabilidad hay de que haya alguno que se comprometa.

En definitiva, esta comunidad tiene muchos usuarios casuales y unos pocos comprometidos, pero el recurso que se genera en esta wiki parece útil.

Ahora pasamos a ver como se distribuye la cantidad de trabajo entre los usuarios, para si ver si los usuarios más comprometidos son los que más trabajo realizan o no.

Para ello nos vamos a centrar en los meses donde hay una comunidad más variada en cuanto al compromiso, es decir, entre los años 2009 y 2010. Como podemos observar en las figuras 6.32, 6.33 y 6.34 que durante el 2009, la mayoría de las ediciones son realizadas por usuarios nuevos y que, en cambio, es a partir de enero de 2010 cuando ya son los usuarios con un mayor compromiso los que toman posesión de la wiki, realizando un mayor número de ediciones. Esto nos puede indicar, la falta de veracidad en las ediciones que estos usuarios novatos realizan sobre el 2009 y que los más experimentados se dedican a corregir durante el año siguiente.

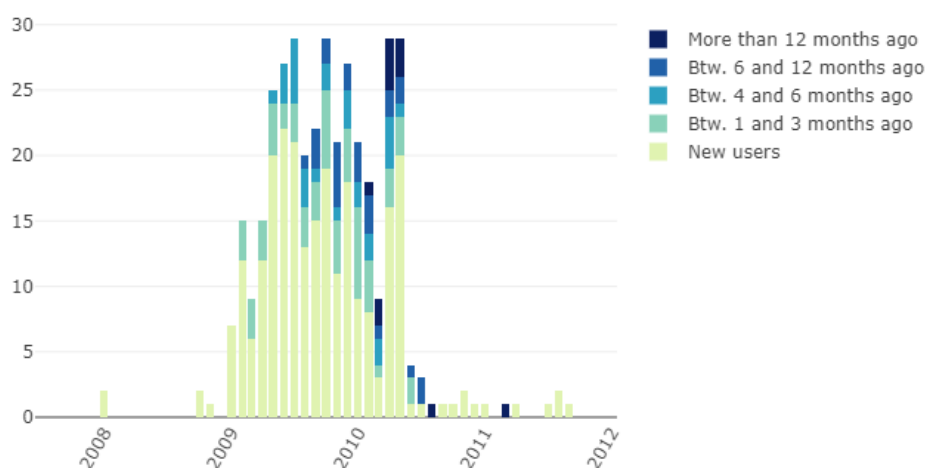


Figura 6.32: Métrica de usuarios activos registrados según su antigüedad, en la wiki de Wikianswer Pets, desde julio de 2007 hasta enero de 2012

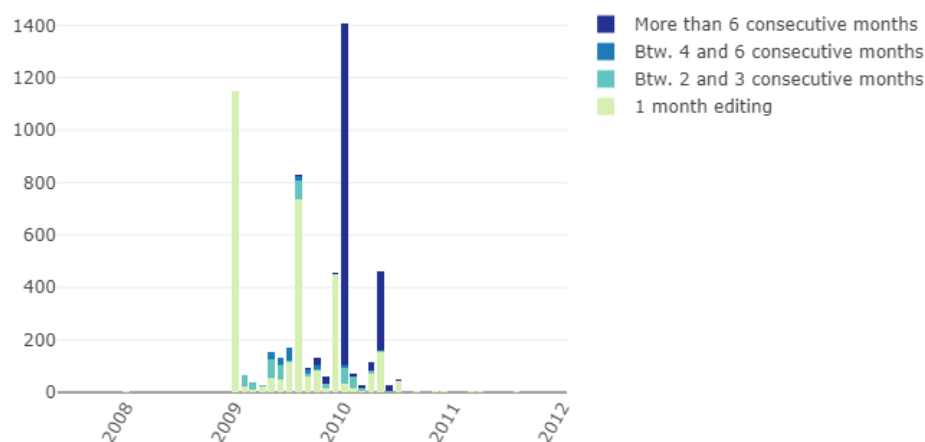


Figura 6.33: Métrica de usuarios activos registrados según su racha actual, en la wiki de Wikianswer Pets, desde julio de 2007 hasta enero de 2012

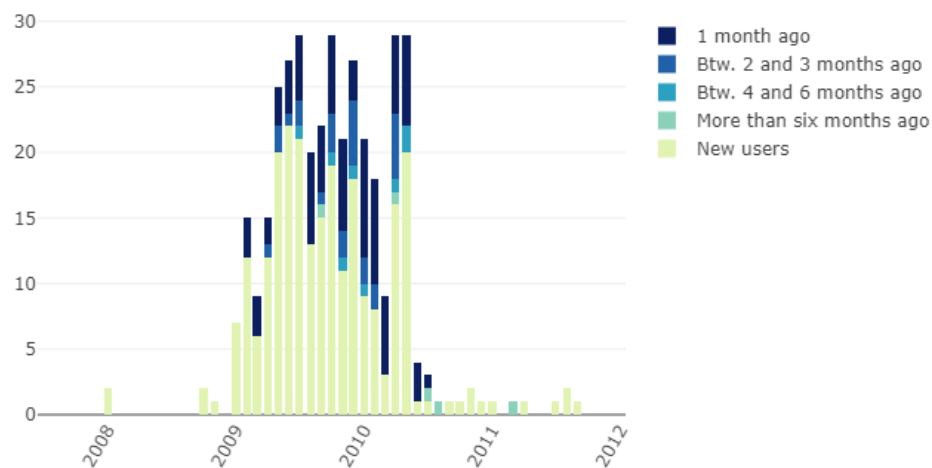


Figura 6.34: Métrica de usuarios activos registrados según su última edición, en la wiki de Wikianswer Pets, desde julio de 2007 hasta enero de 2012

En definitiva, en este caso, los usuarios quemás trabajo realizan son tanto los usuarios muy novatos como los muy experimentados y no hay término medio.

6.4. Análisis sobre ediciones

6.4.1. Análisis con la wiki de Lego(ES)

Para analizar las ediciones dentro de una wiki, contamos con 5 métricas. La wiki elegida para observar fenomenos curiosos es la wiki de Lego en español <https://lego.fandom.com/es/wiki/Portada>. Esta wiki principalmente trata sobre el famoso juego de bloques, pero tambien existen articulos sobre, series de televisión de lego e incluso videojuegos basados en distintas peliculas desarrollados en un mundo de lego.



Figura 6.35: Lego(Es) : Datos de la wiki

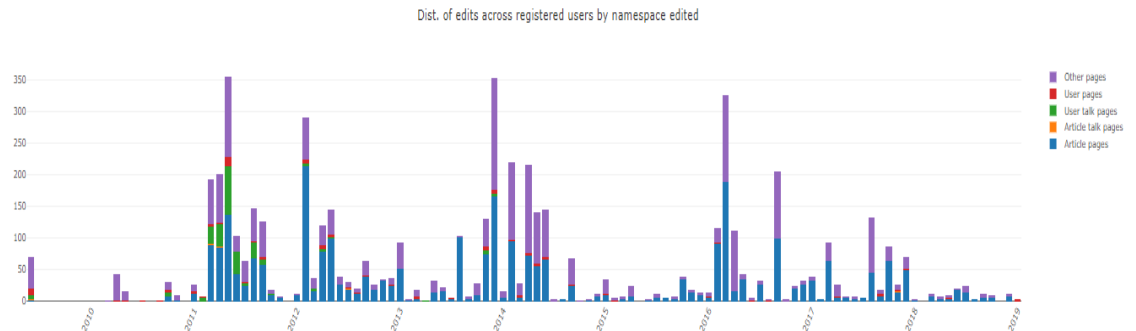


Figura 6.36: Métrica de dist. ediciones entre usuarios activos registrados según el tipo de página editada, en la wiki de Lego

Esta wiki fue creada en Julio de 2009 y el primer mes hubo, como se observa en la imagen 6.36, alrededor de 50 ediciones. Nos sorprende que al crear una wiki existen este número tan alto de ediciones en artículos, esto puede deberse a que se quiere dar calidad a estos artículos ya que son de los primeros o por otro lado podrían ser microediciones para corregir erratas con

el proposito de que la gente que está buscando información interesantes sobre un tema le aparezcan estos artículos de calidad y se queden en la wiki, ya sea como un usuario que solo consulta o en algún momento se puede animar y empezar a editar. Desde la creación de esta wiki, estuvo inactiva, sin tener ninguna edicion hasta abril de 2010, que volvió a tener ediciones durante dos meses consecutivos, en estos dos meses no hubo prácticamente ediciones en articulos, ni en páginas de discusión y casi todas las ediciones se centran en el resto de páginas.

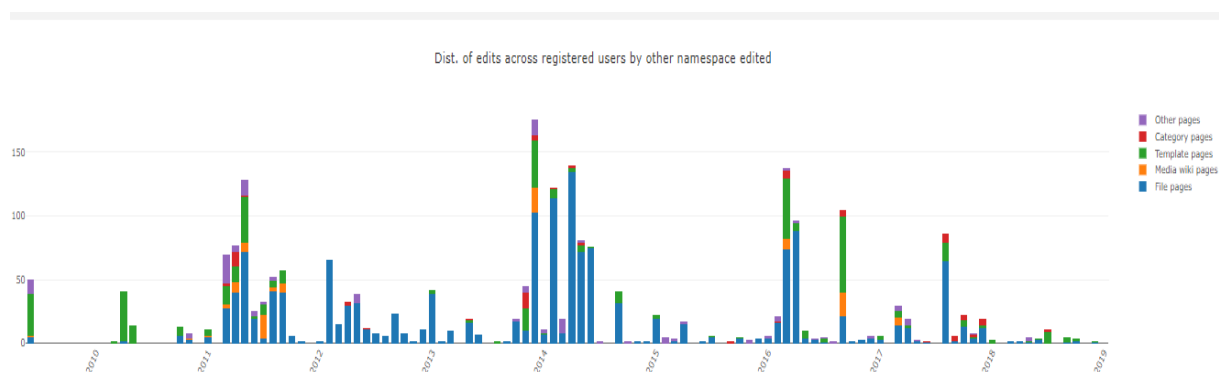


Figura 6.37: Métrica de dist. ediciones entre usuarios activos registrados según el tipo de página editada tomando las categorías de menos importancia, en la wiki de Lego

En la imagen superior (6.37) desglosamos la categoría que anteriormente acumulaba los tipo de páginas no filtradas y nos muestra que al principio de la wiki el tipo de página con más ediciones son las template pages, esto se debe a que al principio de la wiki los editores más activos crean los tipos de modelo de páginas que van a ser utilizados en esta wiki.

A partir de 2011, cuando la wiki comienza arrancar, las páginas que reciben más ediciones, son las 'File pages', en esta wiki que trata en su mayoría sobre figuras lego, es normal que este tipo de páginas sean las más editadas ya que se usan para describir distintas imágenes que en muchos casos, son distintas construcciones lego que distintos usuarios publican y posteriormente explican.

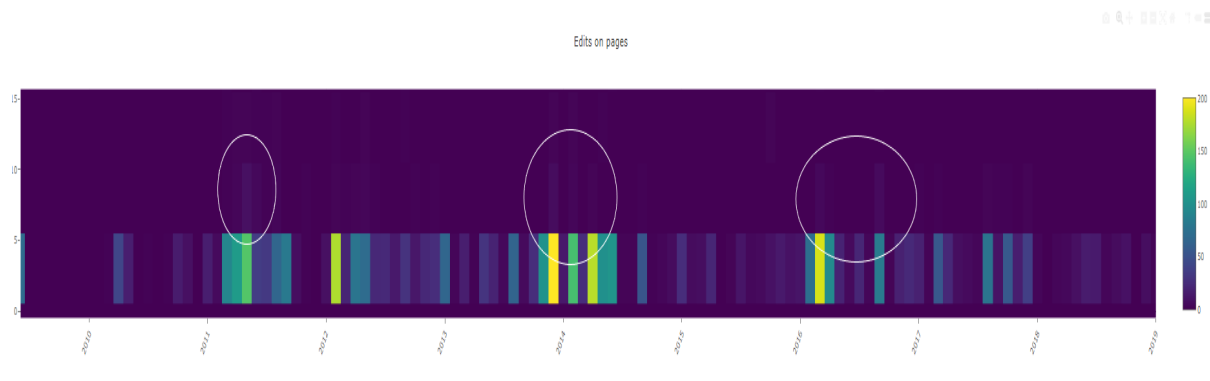


Figura 6.38: Métrica de dist. ediciones en páginas, en la wiki de Lego

En este heatmap (6.38) se observa el número de páginas (eje Z) que tienen el mismo número de ediciones (eje Y) por mes. Después de mayo de 2011 donde hubo un subida clara de ediciones, regresó otro tiempo de inactividad, menor que el anterior y tras este tiempo la wiki empezó a coger soltura e incluso en un mismo mes alrededor de 200 páginas llegaron a recibir un mínimo de una edición y un máximo de 5.

Cabe resaltar, como viene redondeado en la imagen 6.38 que aunque han sido únicamente en 5 meses, ha habido unas pocas páginas que han recibido entre 5 y 10 ediciones, a lo que nosotros deducimos, que puede ser porque estas páginas han tenido mucha repercusión y muchos usuarios han editado para mejorarlo o que ha habido un tema que ha generado mucha controversia y se ha tratado en las páginas de discusión.

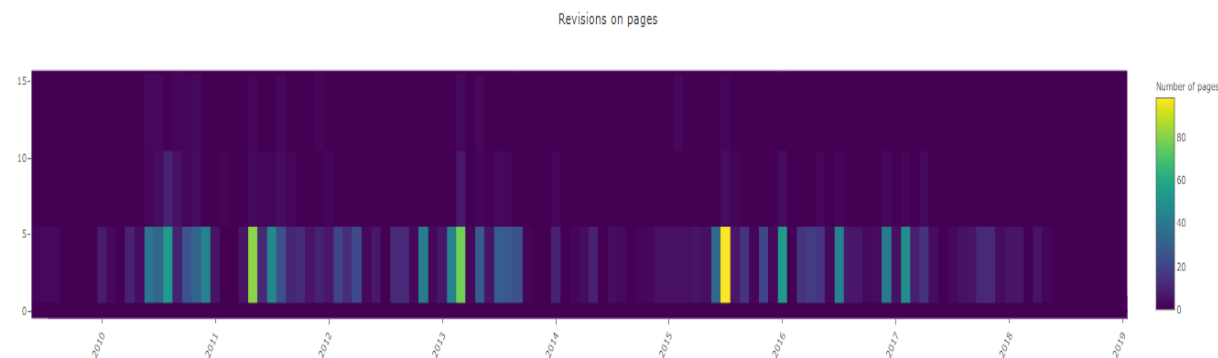


Figura 6.39: Métrica de dist. revisiones en páginas, en la wiki de Lego

La imagen superior (6.39), nos muestra que desde el principio de la wiki ha existido un mínimo de una revisión en cada mes, que ocurra esto es muy positivo para la wiki porque es un indicador de que las páginas no se crean y se dejan en el olvido si no que son revisadas para añadir o eliminar contenido

o corregir erratas. Al comparar esto con la barra de color de las ediciones (6.38), vemos que muchas mas páginas reciben ediciones que revisiones, esto se debe a que al crear una página nueva, esto suma como edición pero no como revisión. Hay muchas páginas que son creadas y nunca mas editadas, puede deberse a que estas páginas se quedan en el olvido o por otro lado son páginas principales de la wiki creadas por los usuarios más fieles y no necesitan ser modificadas, pero podrían serlo en el caso en que en otros meses se consiguiese más información. No nos parece reseñable que tanto en las revisiones como en las ediciones muchas páginas reciben pocas revisiones y muy pocas páginas reciben muchas revisiones, es decir, la masa de páginas se concentra en la parte baja del heatmap, donde las ediciones y revisiones están entre 1 y 5 ya que son métricas que exceptuando la creación son muy parecidas.

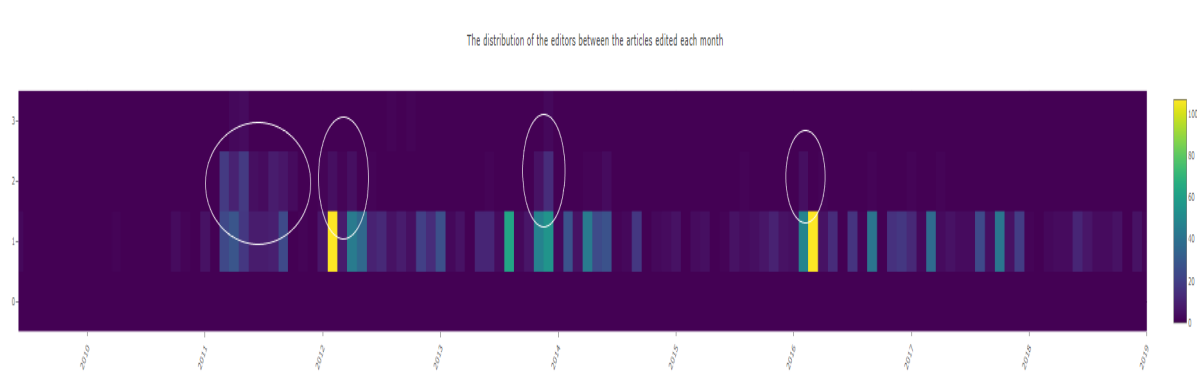


Figura 6.40: Métrica de dist. editores entre artículos, en la wiki de Lego

Este heatmap (6.40) nos muestra los artículos en común editados en un determinado mes por distintos usuarios. En esta wiki muchos usuarios editan artículos individualmente no existe una gran masa de usuarios muy comprometidos que serian los fundadores de la wiki, que esten concentrados en editar los mismos artículos en los mismos meses, ya sea para mejorarlos, eliminar distintas erratas o que sean artículos que crean mucha controversia. En los círculos que aparecen en la imagen se muestra los meses donde varios usuarios si comparten las ediciones del mismo artículo, queremos resaltar que en el círculo entre marzo de 2011 y octubre de 2011, dos editores comparten mas de 15 artículos, esto seria una muestra de dos usuarios muy comprometidos o fundadores que están inteniendo dar calidad a los artículos, que como vimos en la imagen (6.36), además estos meses coinciden que el tipo de página mas editado fueron los artículos, seguramente el proposito de estos editores es poder levantar la wiki, que como se ve en los siguientes meses lo consiguen, porque aumenta la cantidad de ediciones y de editores (6.38).

6.4.1.1. Conclusión:

El objetivo de este análisis es poder observar fenómenos curiosos que nos revelan las ediciones, según su fluctuación tanto en tamaño como en tipo, ahora relacionándolo con hechos importantes en el mundo real que acompaña a esta wiki. Nos gustaría exponer los hechos que más nos ha llamado la atención de esta wiki, por ejemplo, como se ve en la imagen (6.36), hasta 2012 había muchas ediciones en las páginas de usuario de discusión, lo que nosotros relacionamos a que los primeros usuarios que querían levantar la wiki, hablaban muchos entre ellos, para organizar la wiki y repartirse el trabajo. Por eso en estos meses como hemos dicho anteriormente (6.40), hay dos usuarios que editan los mismos artículos el mismo mes.

De esta wiki también nos parece curioso (6.36), que en la mayoría de las wikis el tipo de página más editada son los artículos con una diferencia muy superior a las demás, y esta wiki es un buen ejemplo de otro tipo de página mayoritario como serían las 'File pages' que como hemos explicado antes se debe al contenido de la wiki que este tipo de página sea principal. En ocasiones nos es necesario buscar información sobre algunos fenómenos que se dan en el tema del que trata la wiki para entender ciertos picos y subidas tanto de ediciones como de participación, por ejemplo, en la imagen (6.38) aparecen 4 meses amarillos, donde han recibido ediciones más de 200 páginas, investigando sobre estos cuatro meses, vemos que en todos los casos se debe a noticias importantes que se han dado en el mundo de Lego.

En el primer mes amarillo (Febrero, 2012), se amplió el contrato entre Lego y Lucasfilm, en el segundo caso (Diciembre, 2013), hubo una exposición en Madrid de edificios emblemáticos contruidos con Lego, en el tercer caso (Abril, 2014), se estrenó Lego la película y en el último caso (Marzo, 2016), se publicó un trailer del próximo videojuego de Lego, además de salir a la luz los resultados anuales de la empresa. Para terminar, nos parece muy interesante que el análisis de las wikis en general nos da distinta información sobre sucesos que han pasado en el mundo de cada wiki, y en este caso como hemos hecho investigando poder descubrir a que se debe estos picos y si han colaborado a la supervivencia de la propia wiki.

6.4.2. Análisis con la wiki de Juego de Tronos (GoT)

Vamos a continuar con el análisis sobre ediciones, apoyándonos sobre la wiki de Juego de Tronos https://gameofthrones.fandom.com/wiki/Game_of_Thrones_Wiki, esta wiki trata sobre la famosa serie de HBO, Juego de Tronos, que esta basada en las novelas Canción de hielo y fuego de George R.R. Martin, compuesta por 73 episodios a lo largo de 8 temporadas desde 2011 hasta 2019. Como se ve en la siguiente imagen (6.41), esta wiki es mucho mayor que la de Lego (6.35) ya que posee 129.000 ediciones respecto a las 7.000 que tiene la de Lego, lo mismo pasa con usuarios, artículos y páginas.



Figura 6.41: Datos de GOT

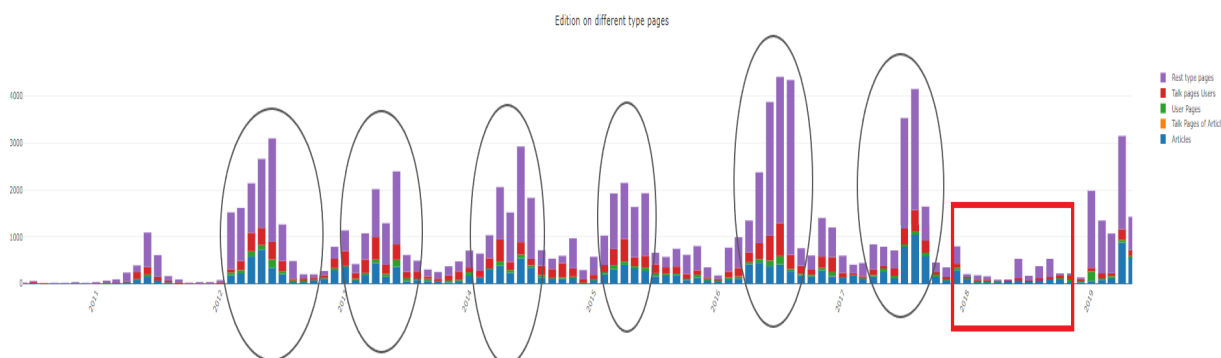


Figura 6.42: Métrica de dist. ediciones entre usuarios activos registrados según el tipo de página editada, en la wiki de GOT

La imagen 6.42, nos muestra que la wiki de Juego de Tronos fue creada a mediados de 2010, pero se mantuvo prácticamente muerta, hasta 2011 cuando empezó a revivir seguramente porque fue la fecha en la se estrenó la serie. Hubo también una pequeña franja de meses, de finales de 2011 hasta febrero de 2012 donde la wiki volvió a decaer. Tras esto como se puede ver en los círculos de la imagen, todos los años se repite el mismo patrón, al principio de año y al final del mismo hay muchas menos ediciones que a mitad esto se debe a que Juego de Tronos excepto las ultimas dos temporadas siempre se ha emitido entre abril y junio. También se observa un descenso de la wiki, entre 2018 y 2019 (cuadro rojo), debido a que este año no se emitió ninguna temporada. Nos resulta curioso que en todos los meses prácticamente desde que la wiki revivió ha existido siempre ediciones en las páginas de discusión de usuario, pensamos que esto es indicativo de una buena comunicación entre usuario ya sea para dar la bienvenida a un usuario nuevo, amonestarle o comentar cualquier duda perteneciente a la wiki o incluso a la serie.

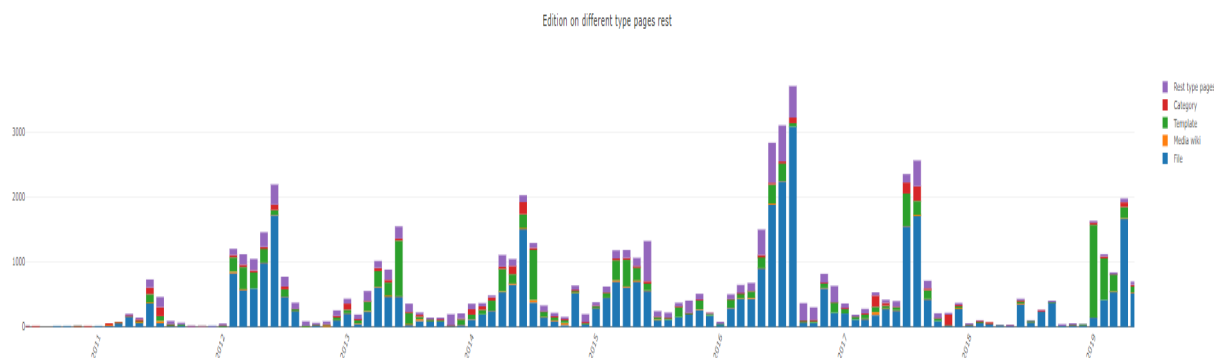


Figura 6.43: Métrica de dist. ediciones entre usuarios activos registrados según el tipo de página editada tomando las categorías de menos importancia, en la wiki de GOT

En la imagen (6.42), se puede observar como la categoría que recibe más ediciones a lo largo de toda la wiki es la que nosotros hemos denominado como rest category, que engloba todas las páginas que no son artículos, páginas de usuario o páginas de discusión de las dos anteriores. Esto no ha resultado curioso, ya que la página predominante en la mayoría de las wikis son los artículos, por eso hemos granulado esta categoría rest, en la imagen (6.43), dividiendo esta, en file, category, media wiki, template y de nuevo rest que tendría todas las páginas que no son las nombradas anteriormente ni las que apareceren la imagen del principio (6.42). De esto podemos deducir que en esta wiki aparte de los artículos que siempre predominan, los archivos y las plantillas son dos tipo de páginas muy editadas. En esta wiki como en la anterior analizada (lego) las “file pages” son de los tipo de páginas que más ediciones recibe de todas, ya que en esta es donde se almacenan los videos e imágenes y se dialoga sobre ellas, tratandose de una wiki de un serie de televisión, es normal que los usuarios publiquen distintos clips de video e imagenes de la serie para resolver dudas ,comentar distintas escenas o analizar lo que podria pasar en la siguiente entrega de la serie.

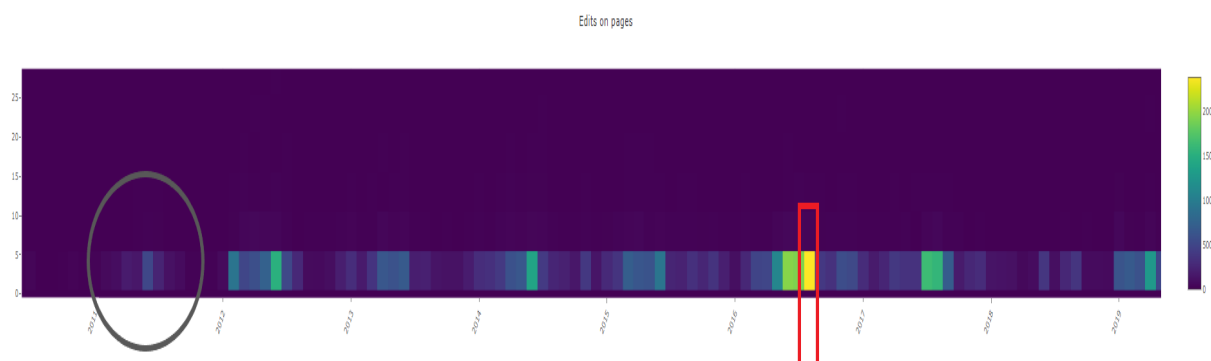


Figura 6.44: Métrica de dist. ediciones en páginas, en la wiki de GOT

En esta imagen (6.44) podemos observar el número de páginas que tiene Y ediciones, desde el principio prácticamente han habido ediciones en todos los meses, aunque como hemos comentado antes, no empezó a arrancar hasta 2011 (marcado con un círculo gris). Si nos fijamos en los valores que abarca el eje Z en la barra de color, vemos como esta wiki es mucho más extensa que otras con hasta 2000 páginas que acumulan ediciones en el mismo mes, por ejemplo en comparación a Lego (6.38), que tenía como cota superior 200 páginas. Visualizando este heatmap, en general, vemos como muestra el mismo fenómeno que la primera imagen (6.42), entre cada año se ve una franja que va aumentando de ediciones, es decir, toma cada vez un color más azulado incluso llegando en algunos casos al verde y vuelve a disminuir al color morado coincidiendo con los meses de no emisión.

Hay que destacar un mes (recuadrado en rojo) que existen muchas más ediciones que los demás meses, buscando información observamos que se debe a que en ese mes publicaron la noticia que la octava temporada sería la última y esto pudo crear mucha inquietud en la wiki.

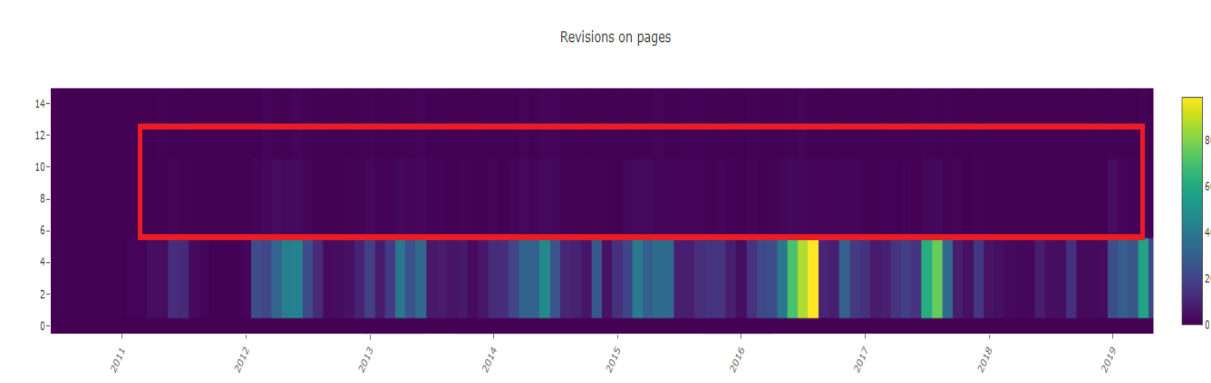


Figura 6.45: Métrica de dist. revisiones en páginas, en la wiki de GOT

En la imagen 6.45, observamos como los meses que más revisiones reciben coinciden en su mayoría con las ediciones, ya que en muchos casos las revisiones son contempladas como ediciones, aunque el número máximo de páginas revisadas (800) es mucho menor que páginas editadas (2000).

En este heatmap (6.44) y en el anterior (6.45), vemos como la mayoría de páginas reciben entre 1 y 5 ediciones y revisiones, es decir, el grueso del heatmap se concentra en la parte inferior, en el caso de las revisiones, se observa mejor como en algunos casos unas pocas páginas reciben entre 5 y 10 revisiones, estas páginas son mucho mas fiables que otras debido a que se ha cuidado mucho mas su información, lo mismo pasa en el heatmap de ediciones pero se observa mucho peor porque el rango de color hasta llegar a 2000, lo oculta en un morado muy oscuro pero como en revisiones llega a 800 estas pocas páginas revisadas las visualiza mejor (recuadro rojo). Como ya comentamos anteriormente que haya una gran cantidad de revisiones es un indicador positivo para una wiki mejorando su calidad, ya sea añadiendo información, eliminando la sobrante o corrigiendo erratas.

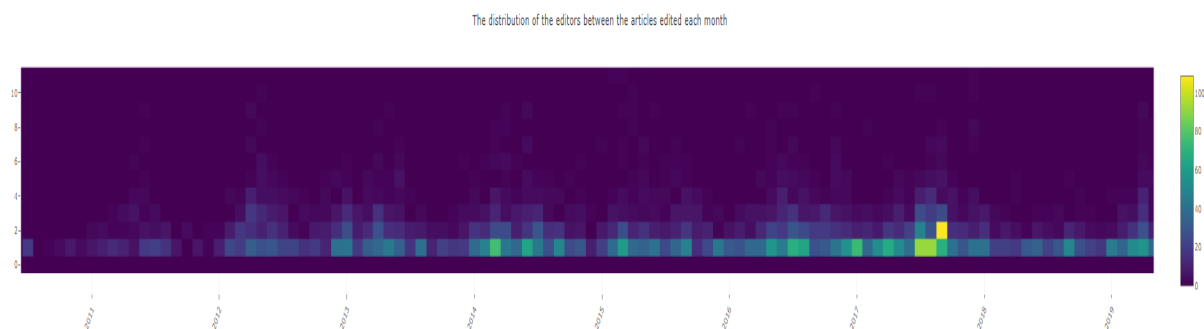


Figura 6.46: Métrica de dist. editores entre artículos, en la wiki de GOT

Este heatmap (6.49) muestra el número de artículos que son editados por los mismos usuarios, por eso en su parte inferior tiene una franja casi completa, ya que representa el número máximo de ediciones sobre artículos de un único usuario. En esta wiki hay editores que son muy fieles a la wiki ya que llegan a editar en un mismo mes hasta 100 artículos, esto es un buen indicador de compromiso con la wiki. Este heatmap se extiende superiormente mucho mas que el de Lego (6.40), ya que al tener muchos más usuarios y muy implicados en la wiki, coinciden en muchos artículos, como se ve en ocasiones existen hasta 4 personas que editan en un mes los mismos 5-6 artículos, estos son artículos de gran repercusión ya que son editados en muchos casos para reforzar la información introducida por otro editor o que consideran que algo no está bien y lo borran para evitar vandalismo en las wikis, eliminando información que es falsa o confusa. Estos artículos que son editados por varias personas son de alta calidad debido al trabajo que hacen los distintos

editores por mejorarlos. La temporada que más colaboración genera es la octava y sobretodo el anuncio de la fecha de emisión esta como se ve entre junio y septiembre de de 2017.

6.4.2.1. Conclusión

La wiki de Juego de Tronos, visualizada con ediciones nos parece que representa muy bien, los meses de emision o no emisión de la serie incluso los meses en los que aunque haya ninguna emisión, aparecen noticias importantes relacionadas con la serie ya que en las barras se ve claramente los picos que se generan en cuanto a cantidad de ediciones hablamos y en los heatmap, se observa como en muchos casos tenemos colores mucho mas fuertes en cuanto estamos en los meses más activos de la wiki. Se observa muy bien la fluctuación de ediciones asi como, el tamaño de la wiki, ya que si comparasemos todas las gráficas, con las gráficas representadas para Lego, nos damos cuenta como en los Heatmap de Juego de Tronos el eje Z es muy superior al de este. Ahora que la serie ha finalizado no podríamos asegurar si esta wiki va a continuar activa, como ha pasado con la de Harry Potter, o por el contrario empezaran a disminuir las ediciones por meses como paso entre 2018 y 2019, donde no hubo emisiones, hasta desaparecer completamente, dependera mucho de los usuarios que están muy arraigados a la wiki si siguen editando o se cansan al no tener nueva información que publicar y abandonan.

6.5. Análisis con Factoides

6.5.0.1. Análisis con la wiki World Witches Series

Esta wiki trata sobre un proyecto multimedia adaptado en dos series de novelas ligeras, dos series de manga, un OVA, una película y una serie de anime por televisión con dos temporadas. La historia gira en torno a las Witches, chicas con poderes mágicos. El OVA fue lanzado el 1 de enero de 2007 como una vista previa del anime. El anime se estrenó el 3 de julio de 2008, concluyendo el 18 de septiembre. La confirmación oficial de la segunda temporada del anime fue anunciado el 27 de febrero de 2009.



Figura 6.47: Datos de World Witches Series

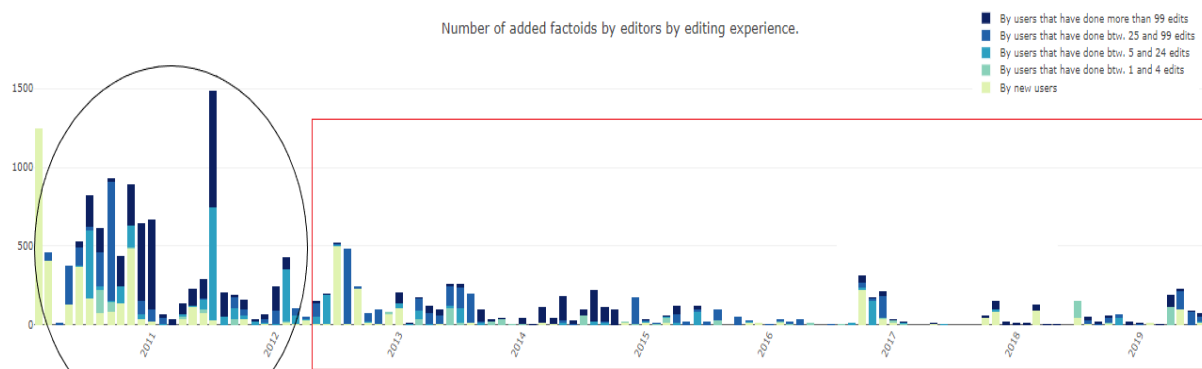


Figura 6.48: Métrica de número de factoides añadidos por editores según su experiencia editando, en la wiki de Word Witches Series

En esta imagen observamos los factoides añadidos, por distintos editores según su experiencia. Hasta prácticamente el cuarto mes de la wiki, los factoides añadidos son por usuarios poco experimentados dentro de la wiki, ya que esta la van adquiriendo con el paso de los meses. A partir de este cuarto mes, ya se pasa de prácticamente tener solo editores nuevos o que tienen como mucho 25 ediciones a diferenciarse más el segmento con editores que hacen entre 25 y 99 ediciones y más de 99.

Los meses que se encuentran dentro del círculo se observa como respecto a la wiki son meses de mucha actividad añadiendo factoides, hay un mes que prácticamente llega a los 1500 y la mitad de ellos son añadidos por editores muy expertos, nos parece curioso como antes de este mes, existen 5 meses consecutivos sin casi actividad, solo unos pocos factoides añadidos principalmente por editores expertos.

La parte que está dentro del cuadro rojo, se mantiene durante toda la wiki entre 0 y 500 factoides añadidos, hasta prácticamente 2015 todo lo que se añade es por editores muy experimentados, después de 2015 hay unos

meses sin casi actividad que luego resurge por la aparición de editores sin casi experiencia que empieza a añadir factoides, estos editores seguramente pasen a ser experimentados en meses próximos.

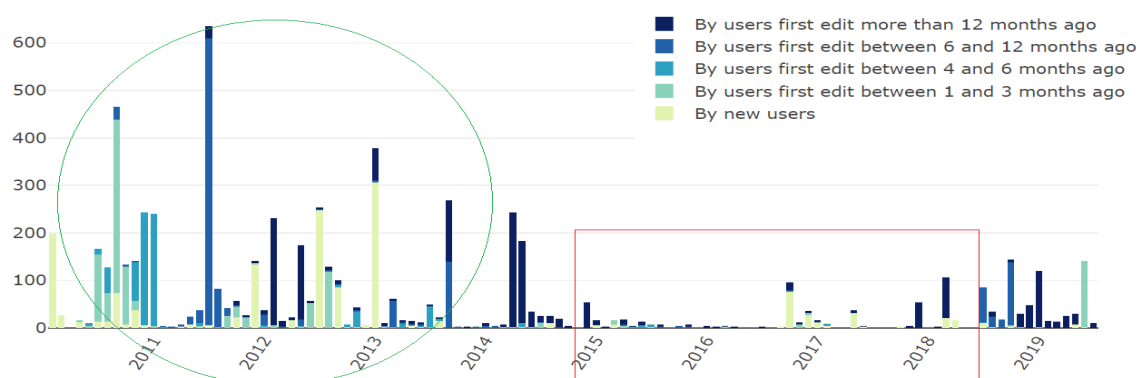


Figura 6.49: Métrica de número de factoides eliminados por editores según su primera edición, en la wiki de Word Witches Series

En la imagen superior, se observan los factores borrados por editores según el tiempo que ha pasado desde su primera edición, como en la imagen anterior (6.48), en el círculo verde están incluidos los meses en los que más factoides se borran, en algunos casos hasta 630, como se puede ver, los editores en los primeros meses no son del tipo de editorexporadico que edita una vez y no vuelve a editar en ningún momento, sino que a medida que pasan los meses, los editores van incrementando de categoría. Los meses fuera del círculo se mantienen constantes entre 40 y 200 factoides eliminados. La parte central, en el recuadro rojo, todo lo que se borra es por editores que llevan mucho tiempo en la wiki, no existen casi editores nuevos, estos editores son los que mantienen viva la wiki aunque no eliminen casi factoides. Los meses finales posteriores al recuadro rojo nos parecen reseñables ya que aumenta la eliminación de factoides por editores muy comprometidos en la wiki, después vuelve a reducirse y a final de 2019 se manifiesta un mes con mucha eliminación por editores muy novatos.

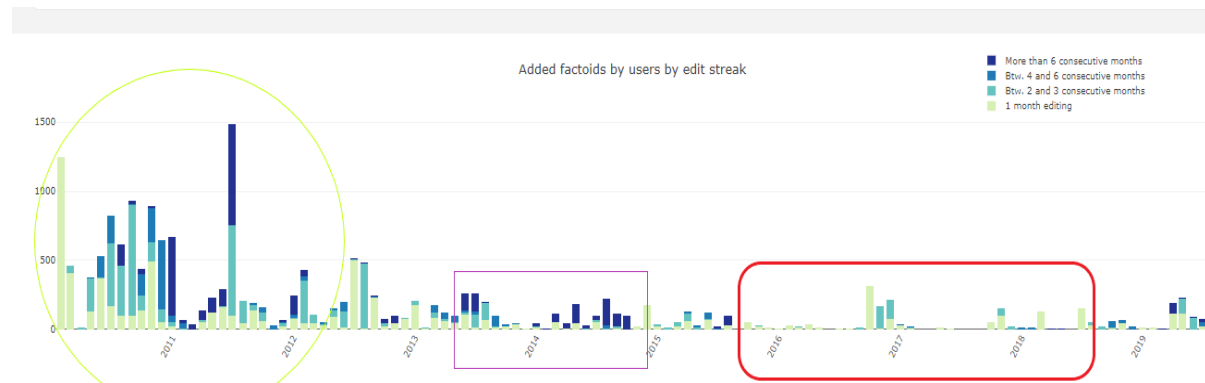


Figura 6.50: Métrica de número de factoides añadidos por editores según su racha actual, en la wiki de Word Witches Series

En esta imagen, se observan los factores añadidos por editores según los meses consecutivos que llevan editando la wiki, en el círculo azul, como en la primera imagen (6.48), son los meses donde más factoides se añaden, gracias a esta métrica vemos perfectamente como los editores, al principio, intentan levantar la wiki, porque a medida que pasan los meses los editores no solo no abandonan la wiki, sino que se van comprometiendo, editando casi todos los meses, por eso en los meses finales del círculo una buena parte de cada barra son editores que llevan editando entre 3 y 6 meses y más de 6 meses. En la parte del recuadro morado, hay una bajada en el número de factoides añadidos, los que se añaden son por usuarios que llevan más de 6 meses editando, que al ser los más fieles a la wiki, intentan que esta no desaparezca como pasa con otras. Pero como se ve en el recuadro rojo estos editores comprometidos también se cansan de luchar por la wiki y abandonan y durante todos estos meses son editores nuevos y en algún caso editores que llevan 2 y 3 meses consecutivos editando son los que añaden distintos factoides.

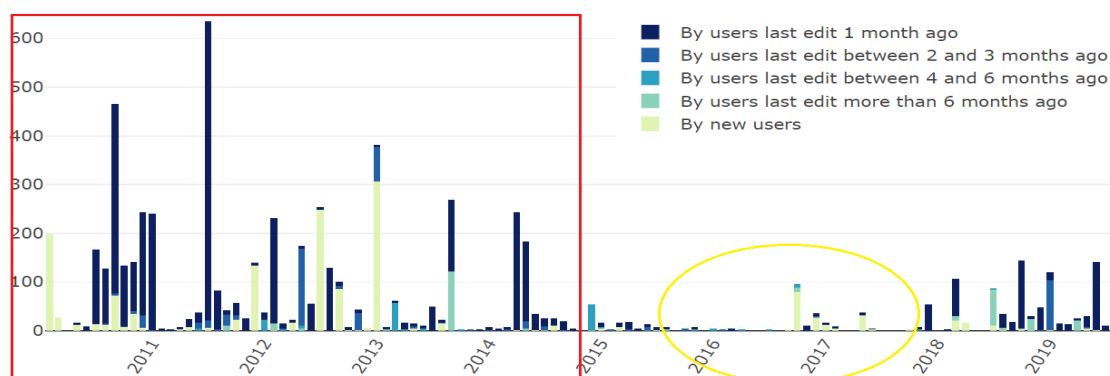


Figura 6.51: Métrica de número de factoides eliminados por editores según su última edición, en la wiki de Word Witches Series

La imagen 6.51 nos muestra los factoides borrados por editores según el tiempo que ha pasado desde su última edición en la wiki, como se ve en el recuadro rojo exceptuando algunos meses como entre septiembre y mayo de 2011 y principios y finales de 2014 el número de factoides borrados se mantiene entre 250 y 600, y en todas las barras la categoría mayoritaria son editores que editaron el último mes, esto puede ser por dos razones, porque hay editores que editan prácticamente todos los meses o porque se van reciclando, es decir, dejan de editar y son los nuevos usuarios los que ahora se transforman en editores que editaron el mes pasado, gracias a la gráfica de edit streak (6.50), sabemos que no se reciclan sino que son los mismos usuarios los que cada mes pertenecen a la categoría de usuarios que han editado el último mes.

También nos gustaría resaltar que hay 3 meses en 2012 donde la mayoría de factoides eliminados son por usuarios nuevos, cosa que no cuadra con el tipo de usuario modelo que elimina factoides en esa wiki que es un usuario comprometido, que edita cada mes.

En el círculo amarillo son los meses en los que menos se eliminan factoides, en los primeros meses son mantenidos por usuarios comprometidos que van abandonando y a principios de 2017 vuelve a crecer esta eliminación por usuarios nuevos.

Al final de la wiki vuelve a aumentar esta eliminación hasta casi alcanzar los 150 factoides eliminados por mes y vuelven a ser los usuarios comprometidos que editan cada mes, los que principalmente eliminan.

Conclusiones y Trabajo Futuro

7.1. Conclusiones

El producto final de este trabajo de fin de grado, tal y como nos proponemos en la sección de Objetivos en el capítulo 1 de esta memoria, es implementar una nueva versión de una herramienta Web existente (WikiChron Classic), que permita el estudio en profundidad de aspectos más complejos en una única wiki.

Entre otros factores, nos centramos en poder observar cómo es la estructura demográfica subyacente de la wiki a lo largo del tiempo, dividiendo a los usuarios activos en clases atendiendo a diversos factores, con las métricas que miden la distribución de usuarios registrados activos:

- Número de ediciones realizadas en la wiki hasta el momento.
- Fecha de la primera edición.
- Fecha de la última edición.
- Número de meses seguidos en los que el usuario contribuye en la wiki.
- Número de meses seguidos en los que el usuario contribuye en la edición de artículos de la wiki.
- Tipo de página en la que los usuarios activos contribuyen.

Esta división en clases, tal y como se observa en los análisis del capítulo 6 de este documento, permite explicar cualitativamente y con mayor claridad las métricas de series temporales disponibles en WikiChron Classic, como el coeficiente de Gini, o cualquiera de los múltiples ratios proporcionados. Además, permiten observar aspectos nuevos que con la versión anterior a WikiChron Monowiki, no era posible discernir, como el poder observar la

evolución temporal de un desgranamiento de la comunidad de usuarios activos atendiendo a distintas clasificaciones.

Además, proporcionamos métricas adicionales que permiten estudiar cómo varía la distribución de la participación a lo largo del tiempo: es el ejemplo de la métrica de % of editors by % of editions, que calcula qué porcentaje de contribuidores llevan a cabo un 50 %, 80 %, 90 % y 99 % de las ediciones totales de la wiki en cada mes, y de las ediciones mensuales, en la versión en relativo de la métrica.

También proporcionamos un conjunto de métricas para medir la distribución de las ediciones según el tipo de usuario. Por tanto, empleando una combinación de métricas de distribución de usuarios registrados activos y métricas de distribución de las ediciones según el tipo de usuario, se puede hacer un diagnóstico de una wiki en el que se analiza cómo de numeroso y de colaborador es cada tipo de usuario a lo largo del tiempo.

Por otra parte, proporcionamos diversas métricas centradas en medir la distribución de la participación en torno a las páginas, en forma de revisiones y ediciones. Estas métricas permiten analizar la evolución del contenido de la wiki, e incluso, en un futuro, determinar valores medios estándares en número de revisiones realizadas para conseguir artículos de calidad, lo que requeriría una investigación más exhaustiva del contenido de la wiki, como la llevada a cabo en Viégas et al. (2004).

Además, hemos añadido un conjunto de métricas centradas en el estudio del conocimiento en la wiki. Las métricas para medir la distribución de factoides añadidos y borrados según el tipo de editor permiten observar qué tipos de usuario añaden o borran más contenido en los artículos de una wiki a lo largo del tiempo. Por otra parte, las métricas centradas en links e imágenes proporcionan una visión de cómo aumenta o disminuye el número de links e imágenes en los artículos de la wiki, permitiendo discernir cómo de rico es el contenido de los artículos. Según nuestro criterio, un artículo que posee un número de links e imágenes es más comunicativo y está mejor documentado que uno sin links ni imágenes.

En definitiva, el aspecto diferencial de WikiChron Monowiki es que permite comparar varios aspectos de una única wiki que suceden al mismo tiempo, en el mismo gráfico. Esto permite descubrir fenómenos que no son fácilmente observables con métricas de menor granularidad.

Por otra parte, tal y como se muestra en el capítulo 3 de este documento, no existen muchas herramientas que permitan el estudio de comunidades colaborativas como las wikis, y las existentes, no permiten efectuar un análisis en profundidad. Con WikiChron Monowiki, creemos que contribuimos a llenar un espacio en blanco, y a impulsar futuras investigaciones en este área de estudio.

7.2. Trabajo futuro

Existen numerosas técnicas y estudios con los que nos gustaría haber podido experimentar, pero no ha sido posible debido a la falta de tiempo, dado que todavía queda mucho por avanzar en el terreno del estudio de wikis.

Mencionamos a continuación, algunas de las líneas de trabajo futuro que son útiles y que se pueden llevar a cabo al hilo del trabajo ya existente.

7.2.1. Visualizar la evolución del texto de los artículos a lo largo del tiempo

Sería interesante incorporar nuevos métodos para la visualización de aspectos de la wiki, como los diagramas de history flow, realizados por Viégas et al. (2004).

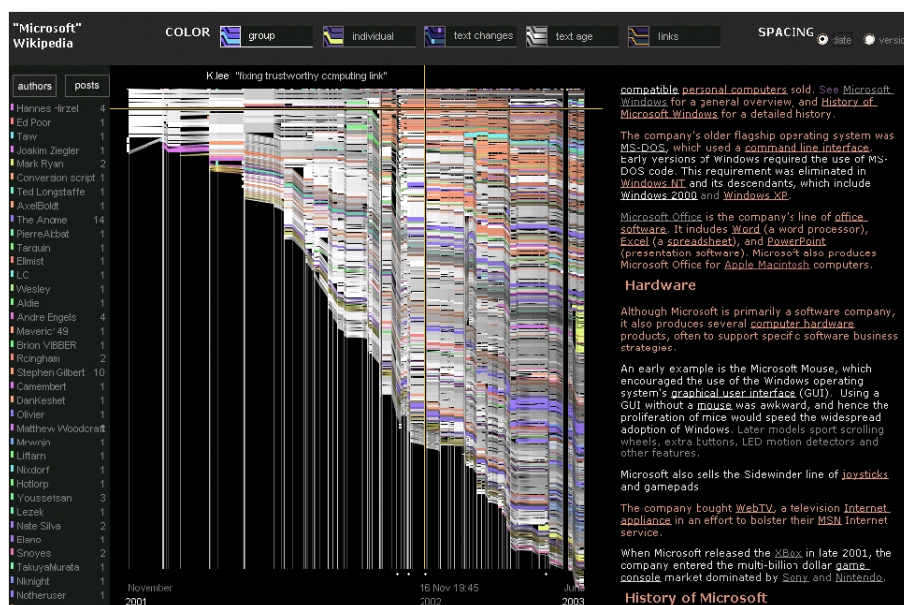


Figura 7.1: Visualización de la interfaz de la herramienta History Flow para la página Microsoft de Wikipedia. A la izquierda, se muestra el contenido de la página, a la derecha, todos los autores que la han editado con un color asignado, y en el centro, la evolución del contenido de la misma.

Como mencionábamos en el marco teórico, estos diagramas (figura 7.1) sirven para observar los cambios que se producen en párrafos del texto de una página, a lo largo del tiempo. En ellos, el texto añadido por cada contribuidor es de un color distinto, y los espacios en negro hacen referencia al borrado de un párrafo que existía en una revisión previa de la página. También se pueden observar cambios de localización de un párrafo a lo largo de la página. Empleando estos diagramas, se podría llevar a cabo un estudio similar al

hecho por estos autores, con el objetivo de determinar qué tipo de cambios sufren las páginas de una wiki a lo largo del tiempo. Esta línea está muy facilitada, dado que el contenido de las ediciones ya está disponible en los csvs tras la modificación del parser que hemos llevado a cabo (ver sección 5.6.1). Esta nueva línea requeriría crear una vista que permitiera al usuario centrarse en un sólo artículo, de todos los correspondientes a una misma wiki.

7.2.2. Refinar la categorización gramatical de las palabras en artículos

En las métricas de factoides de WikiChron Monowiki, necesitamos realizar una distinción entre nombres propios y otras palabras pertenecientes al resto de categorías gramaticales. Para ello, como mencionábamos en la sección 5.6.1, empleamos la librería NLTK de python.

El uso de la librería NLTK de Python para clasificar las palabras del contenido de las ediciones según su categoría gramatical es algo limitado. Esto es debido a que, en Fandom, hay wikis de diversas temáticas, y sus artículos no se encuentran bajo una supervisión y revisión tan estricta como la de proyectos como Wikipedia. Por ello, el contenido de las ediciones contiene palabras mal escritas, o errores gramaticales, que hacen que el NLTK las categorice de forma errónea, y por tanto, un cierto porcentaje de las palabras que se consideran factoides en nuestras métricas, son en realidad palabras mal escritas. Sería interesante elaborar un método sea capaz de tener en cuenta estos errores y de categorizar las palabras correctamente.

7.2.3. Añadir información sobre la intencionalidad de las ediciones utilizando un clasificador

Los autores Yang et al. (2017) han creado un clasificador que se encarga de inferir la intencionalidad del usuario a la hora de llevar a cabo una edición. Estos autores distinguen entre 13 tipos de intención, entre los que se encuentran: clarificación, contra-vandalismo (acción llevada a cabo para evitar un acto de vandalismo), verificación del contenido, etcétera. Empleando este clasificador, se podría incorporar información acerca de la intencionalidad en las ediciones, añadiendo las etiquetas que proporciona el mismo. A partir de ahí, podrían llevarse a cabo nuevas métricas para permitir observar qué intenciones tiene cada tipo de usuario activo.

Conclusions and Future Work

7.1. Conclusions

The final output of this Degree final project, is the adjustment of an existing Web tool, WikiChron Classic, in order to allow a deep study on several and more complex aspects of a single wiki.

Among other aspects, we focus on analyzing how the wiki community behaves, in terms of its underlying demographic structure. We achieve this by dividing the active editors on each month in several classes, according to the following criteria:

- The number of edits that the user has done in the history of the wiki.
- The date in which their first edit was done.
- The date in which their last edit was done.
- The number of months in which the user has consecutively contributed in at least one edit.
- The number of consecutive months in which the user has edited an article.
- The different namespaces the active users have contributed in.

Moreover, for some of these criteria (i.e. the user classes resulting from dividing them according to the total number of edits they have achieved until the moment), we have studied several deeper elements, such as the monthly number and percentage of edits that each group of users has achieved.

This class division, as it was shown in chapter 6 of this document, allows the analyst to give a qualitative explanation and to contribute with qualitative evidence, the results obtained from the quantitative metrics used during

a research, such as the Gini coefficient, or any of the multiple ratios that WikiChron Classic provides.

We also provide a set of metrics which are based on the study of the edits distribution across editors, which allows the analyst to not only tell how numerous is a user group, but also how collaborative the users in this group are. This can be done by combining distribution of active registered users metrics and edits distribution across users metrics.

Moreover, we provide a set of metrics which allow the analyst to study how the distribution of participation evolves for a given percentage of the total editions in the wiki, and the monthly editions in the wiki, by using the % of editors by % of editions on its two versions, monthly and accumulated.

On the other hand, we provide a diverse set of metrics, oriented to showing the distribution of participation (taking into account both revisions and edits) across pages. This metrics allow to study the evolution of content, and might allow, in a near future, to determine standard statistical values to determine the average number of revisions needed to produce a good quality page. This would also require a deeper and more exhaustive investigation, similar to the one carried out in Viégas et al. (2004).

We also provide a set of metrics to study how knowledge evolves in the wiki. Distribution of factoids across registered users metrics can be used to observe which user group adds or deletes more knowledge in the wiki's articles, and links and pictures metrics can be used to observe how rich these articles are. According to our criteria, the more links and pictures an article has, the better documented it is.

Definitely, the differential characteristic of WikiChron Monowiki is its functionality to simultaneously show, in the same graph, several aspects on a unique wiki. This aspect, as we have shown in this Degree final project, allows it to discover several phenomena which are not easy to deduce from less informative metrics, such as ratios or coefficients.

On the other hand, as it was shown in chapter 3 of this document, there are not many tools available which are focused on allowing a deep analysis on wikis, and the existing ones are not as improved at the metric level, as WikiChron Monowiki is. Thus, we have contributed to fill a gap and to promote future investigations in this area.

7.2. Future work

There are several techniques and studies that we would want to experiment with, but due to a lack of time, it was not possible for us. There is still many research to do and many goals to accomplish in the wiki study field. Below, we expose some of the future work lines that are useful and possible to accomplish, using the already done work:

7.2.1. Visualizing text evolution in articles across time

We believe it would be interesting to include graphics that can show new aspects, such as the History Flow diagrams, developed by Viégas et al. (2004).

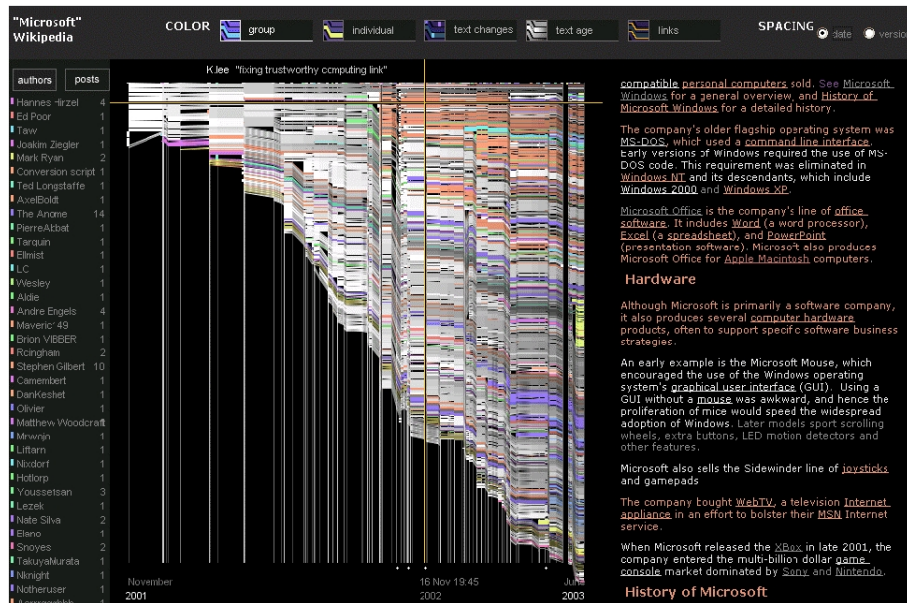


Figure 7.1: History flow visualization for the Wikipedia Microsoft page. We can see the content of the page at right, all the authors that have contributed to it at left, and the visualization of the evolution of the article content in the center.

As we mentioned in the theoretical framework, these diagrams (figure 7.1) are used to observe the changes that text paragraphs from a page suffer across time. We can see that the text added by each author is colored differently, while the black spaces between paragraphs show that some text that was present in a previous revision has been deleted. We can also observe position changes of the paragraphs. By using this diagrams a similar study could be carried out, willing to identify which are the most common changes that wiki articles suffer across time. This new line of research and work has been started by us by adding the content of editions in the wiki csvs (see section 5.6.1). This new work would require the developer to create a new view that would let the user focus on the different articles of a wiki.

7.2.2. Improve part of speech tagging in article words

In order to implement the factoid metrics in WikiChron Monowiki, we need to distinguish among proper nouns and other words belonging to the

rest of the part of speech categories. In order to do so, we have used the NLTK python library, as we said in section 5.6.1.

However, using this library has some limitations, as Fandom wikis have very different purposes, and their articles are not put through a high number of revisions, which means that there are words incorrectly spelled or subject to other grammatical errors. Thus, in our metrics, NLTK is identifying as factoids words that are not really factoids. Thus, we think that it would be interesting to design a method that would take into account the most typical grammatical errors and tag words according to their part of speech category correctly.

7.2.3. Adding intentions information to edits

Authors Yang et al. (2017) have created a classifier that infers the intention of an author when they make a revision on a page. This authors distinguish among 13 different types of intention, such as clarification, counter vandalism or verification of content. By using this classifier, information about the intentions of the author could be added to the edit information, by using the tags that the classifier provides. Then, several metrics to measure the kind of intentions that each type of active user has could be implemented.

Contribuciones individuales

8.1. Laura Sánchez de Rojas Huerta

Al final del curso 2017-2018, comenzamos con una reunión para tomar contacto con nuestros directores, Javier Arroyo y Abel Serrano, que pusieron a nuestra disposición el código de GitHub en el que se encontraba WikiChron Classic, la versión anterior a WikiChron Monowiki, y en la cuál éste se basa.

Al inicio del curso 2018-2019, comenzamos con varias reuniones que tuvieron lugar durante el primer cuatrimestre, en las que nos centramos en tomar contacto con las tecnologías que empleaba WikiChron Classic, que eran desconocidas para mí hasta ese momento. Además, nuestros directores nos sugirieron que contribuyéramos a ampliar WikiChron Classic para entender mejor el funcionamiento de la herramienta, y así lo hicimos. Mi trabajo en esta fase consistió en llevar a cabo algunas de las métricas expuestas en los “issues” del proyecto de WikiChron en GitHub, como la de Usuarios reinidentes, o la de distintos niveles de usuarios activos, que añadí a WikiChron Classic por medio de un pull request.

Además, colaboré en añadir una nueva página en la wiki de WikiChron Classic, para instruir a nuevos usuarios de WikiChron en cómo podían añadir una nueva wiki para analizarla con la herramienta.

Posteriormente comenzamos a crear las métricas que más tarde incorporaríamos en WikiChron Monowiki. En esta fase, mi aportación fue implementar algunas de ellas, en concreto, la métrica de Active editors by experience, Editions by editor experience (versión en absoluto y en relativo), Returning active editors, Active editors by tenure, Active editors in namespaces, % of editors by % of editions, y una métrica más, que mide el cambio en el tamaño absoluto de las diferentes clases de usuarios en función de su experiencia editando. No obstante, esta última métrica no fue añadida a WikiChron Monowiki, dado que la información que aportaba era repetitiva con otras métricas ya creadas.

Más tarde, añadimos estas métricas y las creadas por mis compañeros en WikiChron Classic, empleando el modo de visualización “scatter”, o en líneas, que es el empleado por defecto en esta primera versión de WikiChron. Más adelante, Cuando mis compañeros Adrián y Laura añadieron en el código de WikiChron Classic, el modo de visualización con barras, y una lógica para unir varias submétricas en una misma gráfica, mi trabajo consistió en incorporar algunas de las métricas que habíamos creado, a esta prematura versión de WikiChron Monowiki, organizando y comentando el código. Posteriormente, en una versión más madura de la herramienta, y tras haber incorporado todas las métricas, añadí el modo de visualización en “area chart”, requerido por la métrica de % of editors by % of editions.

Más tarde, mis contribuciones se basaron en recopilar artículos para posteriormente llevar a cabo la parte de investigación teórica de este trabajo de fin de grado, y colaborar con mis compañeros en arreglar pequeños errores que fuimos detectando progresivamente, con ayuda de nuestros directores.

Luego, ayudé a nuestro codirector, Abel Serrano, en la unificación de WikiChron Monowiki con el resto de modos de WikiChron (WikiChron Classic y WikiChron Networks). Una vez unificado con el resto, me encargué de llevar a cabo la reestructuración del código para incluir los distintos tipos de gráficos de la forma en la que está actualmente implementado: usando herencia de clases y métodos, para que sea más eficiente y fácil añadir nuevos gráficos.

Posteriormente, tras ser aceptada esta reestructuración del código por nuestros directores, decidimos ampliar el conjunto de métricas centradas en la distribución de ediciones según el tipo de usuario que las realiza. Me encargué de llevar a cabo estas nuevas métricas, que son las de Edits by users by tenure y Edits by users by the date of the last edit.

Más tarde, comenzamos la nueva línea de trabajo centrada en la elaboración de métricas para el estudio del conocimiento. Después de que Laura De La Fuente modificara el parser que teníamos para obtener los datos del dump de Wikia, me encargué de modificarlo para que incluyera una nueva columna en el csv de las wikis con información sobre los factoides. También elaboré la función que saca los factoides del contenido de las ediciones, usando la librería NLTK de Python.

Con el parser ya en funcionamiento y los csvs de algunas wikis generados, me encargué de elaborar las métricas centradas en factoides y en la Distribución de factoides según el tipo de usuario.

8.2. Laura De La Fuente Lorenzo

Al principio del curso 2018/2019 empezamos con este trabajo, donde en un principio, nuestros directores nos sugirieron realizar algunas contribuciones sencillas a WikiChron tal y como estaba en ese momento para tomar

contacto con la herramienta y con el lenguaje de programación que íbamos a usar. Mis aportaciones en este momento fueron realizar la implementación de las siguientes métricas: “Active users with >4 edit”, “Surviving new editors”, “Returning new editors”, “Active registered users” y “Active anonymous users”. Una vez realizada la implementación hice mi primer pull request de las dos últimas métricas mencionadas al repositorio oficial de WikiChron.

Posteriormente, con la ayuda de nuestros directores, propusimos varias métricas nuevas para desarrollar, algunas de ellas las implementé junto con Adrián Burillo, pero no se incorporaron a nuestro trabajo, ya que tras realizar un análisis de ellas vimos que no eran relevantes. Otras métricas que fuimos propusiendo si que lo eran y dentro de ellas yo he realizado las siguientes: “Dist. of active registered users by edit streak”, “Dist. of active registered users by edit in article streak”, “Dist. of active registered users by editing experience (absolute)”, “Dist. of edits across editors” y “Bytes difference across articles”.

Para todas las métricas propuestas, vimos que eran interesantes estudiarlas para una sola wiki y por eso, entre Adrián Burillo y yo modificamos la selección de los datos para que estas visualizaciones se realizaran sobre una sola wiki.

Tras estudiar un poco las métricas, introdujimos tres nuevos modos de visualización explicados en el punto 5.2.1, mis aportaciones en ese momento fueron incorporar los gráficos de barras y los mapas de calor, este último con ayuda de Laura Sánchez. Por ello, nos vimos obligados a cambiar alguna funcionalidad ya creada en la aplicación como era la descarga de los datos de los gráficos, esto lo modifiqué yo para que se pudieran descargar los datos para cualquier tipo de gráfico.

A finales de mayo, ya teníamos todo el código implementado, y nuestros directores nos propusieron cambiar algunas cosas para mejorar nuestro trabajo y estas modificaciones las realizamos entre los tres.

A principios de verano, además de seguir arreglando errores, decidimos aumentar el trabajo realizado hasta el momento y por ello nos pusimos a implementar nuevas métricas centradas en la ediciones y otras centradas en los factoids. Mis contribuciones en este punto ha sido cambiar los colores de los gráficos de barras, optimizar las métricas que hice previamente y tardaban un tiempo mayor de lo esperado, modificar, con ayuda de Laura Sánchez, el parser para obtener los csvs con los datos de las wikis e implementar las siguientes métricas: “Dist. of edits across registered users by edit streak”, “Total added factoids”, “Total deleted factoids”, “Added factoids per deleted factoids”, “Added factoids per active anonymous users”, “Deleted factoids per active anonymous users”, “Added factoids per new users” y “Deleted factoids per new users” y también ayudé a mi compañera Laura Sánchez a solucionar la métrica “Percentage of registered users making a percentage of edits” en su versión mensual y acumulada. Mientras tanto, también, nuestro director nos dio feedback nuevos sobre el nombre de las métricas y ciertas pequeñas mejoras de la aplicación, de esto me encargué yo.

Por último, se nos dió un último feedback a finales de verano relacionado con las métricas de factoides que nos ocupamos los tres de ello.

8.3. Adrián Burillo Elmaleh

Como han nombrado mis compañeras previamente la búsqueda de este proyecto comenzó al final del curso 2017-2018 donde tuvimos las primeras reuniones antes de comenzar el verano, y nuestros directores nos aconsejaron que durante el verano tomásemos contacto con las tecnologías que íbamos a utilizar en este TFG, principalmente Python, Pandas, Dash y conocer la estructura de la herramienta en ese momento. En septiembre de 2018, tuvimos las primeras reuniones oficiales, que ya incluían las primeras cargas de trabajo, realizando algunas métricas con Laura de la Fuente, para tomar el contacto de como crearlas, publicarlas y utilizar distintas wikis para ver su utilidad, estas primeras métricas no fueron publicadas en Wikichron por su falta de utilidad. Tras estos meses pensando ,discutiendo y realizando distintas métricas escribimos a mitad de un curso un análisis previo al comienzo de la memoria, para ver hasta que punto eran útiles las métricas que habíamos creado y como analizar una wiki, apoyandonos en nuestro conjunto de métricas.

Tras este análisis empezó lo que sería nuestro TFG propiamente dicho, Wikichron Monowiki para ello junto a Laura de la Fuente, modificamos la selección de los datos para poder visualizar varias submétricas sobre una wiki en un gráfico. Tras conseguir esta modificación creamos las primeras métricas que incluyesen varias submétricas y su composición fuese útil. Alrededor de Abril, Abel migró wikichron versión 1, a Wikichron versión 2, y estuve un tiempo arreglando un problema, que no permitía cargar lo csv y por lo tanto no se podía visualizar las gráficas. Tras el paso a esta versión, mis compañeras introdujeron dos métodos nuevos de visualización, HeatMap y Areachart Más tarde junto a Laura de la Fuente, definimos el tipo de métricas que iban a ser viables para ser visualizadas mediante mapas de calor, y estuvimos haciendo distintas pruebas que acabaron trasformandose en métricas útiles, "Edition on pages", "Revision on pages", "Distribution of editors across articles"(junto Laura Sanchez).

Para el análisis final separamos este en tres partes, tendencias estructurales, compromiso de usuarios y ediciones, de este ultimo me encargé yo y para poder explicar mejor el tipo de ediciones, desarrollé dos nuevas métricas "Namespace editedz .Other namespace edited.ambas separan las ediciones según el número de páginas editadas. Además añadí para WikiChron classic la métrica de wikis durmientes, por petición de mi director Javier.

Para la entrega de la memoria yo me encargué, como he dicho anteriormente de desarrollar el análisis de ediciones, la explicacion de WikiChron versión 1, la instalación en Windows y la explicación de la métricas que había

desarrollada durante el año.

Tras la decisión de nuestros directores, de no presentar el TFG en junio por insuficiencia de trabajo, planemos reuniones despues de exámenes para aclarar que ibamos a extender para presentarlo en septiembre. De lo primero que hicimos los tres fue optimizar las métricas porque era un error que nos perseguía desde hace bastante y no podíamos entregar nada que no terminase de ejecutar.

Durante verano, mis compañeras cambiaron la forma de descarga de los Csv para poder ver el contenido de las ediciones y obtener los factoides ,yo trabajé principalmente en desarrollar las métricas de Added y deleted factoids by users by edit streak con la ayuda de Laura de la Fuente siguiendo el esquema implementado por Laura Sanchez, además de crear tres nuevas métricas sobre el conteo de imagenes y links externos e internos.

En cuanto a la memoria final, los tres nos encargamos de aplicar los cambios indicados por nuestro director Javier respecto a la memoria de Mayo, además de añadir un nuevo análisis utilizando las métricas del nuevo apartado de factoides, y amplié el análisis anterior de ediciones.

Bibliografía

*Y así, del mucho leer y del poco dormir,
se le secó el cerebro de manera que vino
a perder el juicio.*

Miguel de Cervantes Saavedra

ALEXA SITE INFORMATION. Alexa: Wikia site information. .

ALLAHO, M. Y. y LEE, W. Analyzing the social ties and structure of contributors in open source software community. En *2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013)*, páginas 56–60. 2013.

ASTA ZELENKAUSKAITE, P. M. Digital libraries and social web: Insights from wikipedia users' activities. *IADIS Multiconference on Computer Science and Information Systems*, 2011.

BATES, B. Yochai benkler. the wealth of networks: How social production transforms markets and freedom. *Journal of Media Economics*, vol. 20, páginas 161–165, 2007.

BAUWENS, M. The political economy of peer production. *Post-Autistic Economics Review*, vol. 37, 2005.

BENKLER, Y. Peer production, the commons, and the future of the firm. *The Harvard Library*, 2017.

BENKLER, Y. y NISSENBAUM, H. Commons-based peer production and virtue. *Journal of Political Philosophy*, vol. 14(4), páginas 394–419, 2006. ISSN 0963-8016.

CHHABRA, A. y IYENGAR, S. R. S. Characterizing the triggering phenomenon in wikipedia. En *Proceedings of the 14th International Symposium on Open Collaboration - OpenSym '18*, página 11. 2018. Exported from <https://app.dimensions.ai> on 2019/05/11.

- CLIFTON, B. *Advanced Web Metrics with Google Analytics*. Sybex, 2012. ISBN 978-1-118-16844-8.
- COFFIN, J. An analysis of open source principles in diverse collaborative communities. *First Monday*, vol. 11(6), 2006. ISSN 13960466.
- CUMMINGS, R. E. *OPEN WRITING: WIKIS, COMMONS-BASED PEER PRODUCTION, AND THE COMPOSITION CLASSROOM*. Versión electrónica, 2006.
- DRAKSHANDEH, Z. Basic google analytics filters for every site. <https://www.bounteous.com/insights/2015/12/10/basic-google-analytics-filters-every-site/>, 2015. [Web; accedido el 10-05-2019].
- ESACADEMIC.COM. Wikia. <https://esacademic.com/dic.nsf/eswiki/1224944>, 2019. [Web; accedido el 10-05-2019].
- FANDOM EXPLORE. Fandom explore. .
- FANDOM FREQUENTLY ASKED QUESTIONS. Fandom frequently asked questions. .
- FANDOM USER RIGHTS. Fandom user rights. .
- FANGZHOU ZHANG, A., WANG, R., BLOHM, E., BUDAK, C., ROBERT, L. y M ROMERO, D. *Participation of New Editors After Times of Shock on Wikipedia*. 13th International AAAI Conference on Web and Social Media, 2019.
- FUSTER MORELL, M. D., WP COORDINATOR); ; SALCEDO, P. D. D. R. M. M. F. C. I. B. M. T. W. A. A. C. A. G. A., J.L.; DE FILIPPI y ROZAS, D. *Value in commons-based peer production*. Versión electrónica, 2015.
- GUPTA, J. N. D. y SUSHIL, S. *Creating Knowledge Based Organizations*. PA: IGI Globa, 2004.
- HAN, P. How to set up goal funnel visualization reports on google analytics. <https://medium.com/analytics-for-humans/how-to-set-up-a-goal-funnel-visualization-report-on-google-analytics-to-identify> 2019. [Web; accedido el 10-05-2019].
- KARPF, D. Blogs, wikipedia, second life, and beyond: From production to produsage, by axel bruns. *Journal of Information Technology & Politics*, vol. 6(1), páginas 81–83, 2009.

- KEYWORTH, P. What can google analytics event tracking be used for? <https://www.hallaminternet.com/using-google-analytics-event-tracking/>, 2017. [Web; accedido el 10-05-2019].
- KHANSA, L., MA, X., LIGINLAL, D. y KIM, S. S. Understanding members' active participation in online question-and-answer communities: A theory and empirical analysis. *Journal of Management Information Systems*, vol. 32(2), páginas 162–203, 2015.
- KITTUR, A., CHI, E., A. PENDLETON, B., SUH, B. y MYTKOWICZ, T. Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. *World Wide Web*, vol. 1, 2006.
- KOSTAKIS, V. Peer governance and wikipedia: Identifying and understanding the problems of wikipedia's governance. *First Monday*, vol. 15(3), 2010. ISSN 13960466.
- LEUF, B. y CUNNINGHAM, W. *The Wiki Way: Quick Collaboration on the Web*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2001. ISBN 0-201-71499-X.
- LIN, Y. y CHEN, Y. Do less active participants make active participants more active? an examination of chinese wikipedia. *Decision Support Systems*, vol. 114, 2018.
- MEDRANO, D. Tableau online tips: Extracts, live connections and cloud data. <https://www.tableau.com/about/blog/2016/4/tableau-online-tips-extracts-live-connections-cloud-data-53351>, 2016. [Web; accedido el 10-05-2019].
- MURRAY, D. G. *Tableau your data!*. Wiley, 2016. ISBN 1119001196.
- NIELSEN, J. The 90-9-1 rule for participation inequality in social media and online communities. *Nielsen Norman Group*, 2006.
- PINK, D. H. The book stops here. *Wired*, 2005.
- RAMAN, M. Wiki technology as a “free” collaborative tool within an organizational setting. *Information Systems Management*, vol. 23(4), páginas 59–66, 2006.
- RETURNING. Returning new editor. Wikimedia Foundation, .
- SALEH, K. 7 essential google analytics reports every marketer must know. <https://searchengineland.com/7-essential-google-analytics-reports-every-marketer-must-know-250412>, 2016. [Web; accedido el 10-05-2019].

- SANGERS, L. Why wikipedia must jettison its anti-elitism. *Larry Sanger blog*, 2004.
- SCHOTT, B. Deletionists. *The New York Times*, 2009.
- SERRANO, A., ARROYO, J. y HASSAN, S. Participation inequality in wikis: A temporal analysis using wikichron. En *OpenSym*, páginas 1–7. 2018a.
- SERRANO, A., ARROYO, J. y HASSAN, S. Webtool for the analysis and visualization of the evolution of wiki online communities. En *26th European Conference on Information Systems: Beyond Digitization - Facets of Socio-Technical Change, ECIS 2018, Portsmouth, UK, June 23-28, 2018*, página 73. 2018b.
- SHAW, A. y M. HILL, B. Laboratories of oligarchy? how the iron law extends to peer production. *Journal of Communication*, vol. 64, 2014.
- SIEFKES, C. The boom of commons-based peer production. https://en.wikipedia.org/wiki/Wikipedia:User_access_levels, 2011. [Web; accedido el 08-05-2019].
- SURVIVING. Surviving new editor. Wikimedia Foundation, .
- ULLRICH, C., BORAU, K., LUO, H., TAN, X., SHEN, L. y SHEN, R. Why web 2.0 is good for learning and for research: principles and prototypes. En *WWW '08: Proceeding of the 17th international conference on World Wide Web*, páginas 705–714. 2008.
- VIÉGAS, F. B., WATTENBERG, M. y DAVE, K. Studying cooperation and conflict between authors with history flow visualizations. En *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '04*, páginas 575–582. ACM, New York, NY, USA, 2004. ISBN 1-58113-702-8.
- VOSS, J. Measuring wikipedia. *Proceedings of ISSI 2005: 10th International Conference of the International Society for Scientometrics and Informetrics*, vol. 1, 2005.
- WAGNER, C. Wiki: A technology for conversational knowledge management and group collaboration. *CAIS*, vol. 13, página 19, 2004.
- WARSCHAUER, M. Audience, authorship, and artifact: The emergent semiotics of web 2.0. *Annual Review of Applied Linguistics*, vol. 27, páginas 1 – 23, 2007.
- WIKIPEDIA DATABASE DOWNLOAD. Wikipedia database download. https://en.wikipedia.org/wiki/Wikipedia:Database_download, . [Web; accedido el 08-05-2019].

- WIKIPEDIA DEVELOPERS. Wikipedia developers. <https://www.mediawiki.org/wiki/Developers>, . [Web; accedido el 08-05-2019].
- WIKIPEDIA REQUEST FOR ADMINSHIP. Wikipedia requests for adminship. https://en.wikipedia.org/wiki/Wikipedia:Requests_for_adminship, . [Web; accedido el 08-05-2019].
- WIKIPEDIA ROLE OF JIMMY WALES. Wikipedia role of jimmy wales. https://en.wikipedia.org/wiki/Wikipedia:Role_of_Jimmy_Wales, . [Web; accedido el 08-05-2019].
- WIKIPEDIA USER ACCESS LEVELS. Wikipedia user access levels. https://en.wikipedia.org/wiki/Wikipedia:User_access_levels, . [Web; accedido el 08-05-2019].
- WIKIPEDIA WATCHLIST. Wikipedia watchlist. <https://en.wikipedia.org/wiki/Help:Watchlist>, . [Web; accedido el 07-05-2019].
- BATTLE FOR WIKIPEDIA'S SOUL, T. The battle for wikipedia's soul. *The Economist*, 2008.
- YANG, A., FAN, H., JING, N., SUN, Y. y ZIPF, A. Temporal analysis on contribution inequality in openstreetmap: A comparative study for four countries. *ISPRS Int. J. Geo-Inf.*, vol. 5, 2016.
- YANG, D., HALFAKER, A., KRAUT, R. y HOVY, E. Identifying semantic edit intentions from revisions in Wikipedia. En *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, páginas 2000–2010. Association for Computational Linguistics, Copenhagen, Denmark, 2017.
- ZHU, H., KRAUT, R. E. y KITUR, A. The impact of membership overlap on the survival of online communities. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, páginas 281–290, 2014.